VOLUME 56 WHOLE NO. 255
NUMBER 2 1944

Psychological Monographs

JOHN F. DASHIELL, *Editor*

Studies in Language Behavior

Iowa university.

UNIVERSITY OF IOWA STUDIES IN PSYCHOLOGY, NO. 24

Edited by WENDELL JOHNSON

I. A Program of Research
By WENDELL JOHNSON

II. The Quantitative Differentiation of Samples of Spoken Language
By HELEN FAIRBANKS

III. The Quantitative Differentiation of Samples of Written Language
By MARY BACHMAN MANN

IV. A Statistical and Comparative Analysis of Individual Written Language Samples
By JOHN W. CHOTLOS

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Publications Office
NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

TABLE OF CONTENTS

STUDIES IN LANGUAGE BEHAVIOR

I. A Program of Research

WENDELL JOHNSON
University of Iowa

The studies by Fairbanks (5), Mann (11) and Chotlos (4), which are presented in this issue of *The Psychological Monographs,* constitute the beginnings of a program of research in language behavior. They have been completed in the order named. The present paper is intended as an introduction to them, and to the general program which they represent.

The importance of language and of symbolization generally, as a distinctly human form of behavior and as a basic factor in personal and social problems, is generally recognized (9, 10, 12, 20). The effective scientific investigation of such behavior, however, depends upon the development of highly reliable and differentiating measures, by means of which specified aspects of language behavior might be systematically observed in relation to one another and to other variables. With such measures, significant testable hypotheses can be formulated and checked, and a body of dependable information can be accumulated.

SPECIFIC OBJECTIVES

The proposed program of research is designed to:

1. Develop reliable and differentiating measures of specified aspects of language behavior.
2. Determine the degree to which the resulting measures are intercorrelated.
3. Determine the degree of correlation between these measures and such other pertinent variables as those involved in environmental influences, physiological conditions, intelligence, and personality adjustment.
4. Apply the measures to a comprehensive investigation of language development.
5. Determine the degree to which language behavior, as measured, is modifiable under specified conditions.
6. Determine the degree to which modification in language behavior is associated with modifications in other aspects of behavior or adjustment.
7. Indicate the normal characteristics of language development and language behavior, and the varieties of disorder or abnormality in such behavior, in terms of the measures used.

Types of Language Measures to Be Investigated

No attempt will be made here to present a review of the theoretical and experimental literature dealing with the problems with which this program is concerned. It is sufficient to say that previous work in the field has suggested many of the procedures to be employed, and that others have been suggested by preliminary research carried out by the writer, or under his direction. A comprehensive review of language behavior studies has been published by Sanford (12). The following types of language measures are to be investigated:

Type-token ratio (TTR). This is a measure of vocabulary "flexibility" or variability, designed to indicate certain aspects of language adequacy. It expresses the ratio of different words (types) to total words (tokens) in a given language sample. If in speaking 100 words (tokens) an individual uses 64 different words (types), his TTR would be .64. In order to develop the most highly reliable and differentiating form of the TTR, it is to be computed for

given language samples in the following various ways:

a. For all words spoken or written by a given individual, or in a given language sample, and separately for words representing the various grammatical categories; for words in different frequency categories—for example, the 500 most frequently used words, the next 500 most frequently used words, etc., as determined by the published word-counts of Thorndike (17), Horn (6), and others, or by the word-counts to be derived from the present investigations; etc.

b. With varying statistical or mathematical procedures, thus:

The over-all TTR, as computed for an entire language sample. TTR's for samples of different magnitudes are not directly comparable because of the tendency for the TTR to vary inversely with size of sample. A knowledge of the precise character of this inverse relationship might make it possible to compare directly TTR's for samples differing in length, by means of a correction table. The feasibility of constructing such a table is to be investigated. The study by Chotlos (4) throws considerable light on this problem.

The mean segmental TTR. TTR's for samples of different magnitudes can be made comparable by dividing each sample into like-sized segments of, say, 100 words each, computing the TTR for each segment and then averaging the segmental TTR's for each sample. It can be safely assumed that such segmental TTR's are directly comparable, so long as they represent segments of equal size, and that means of such segmental TTR's are also directly comparable. Results obtained by using segments of different magnitudes—as 100-word segments, 500-word segments, etc.—are to be compared, in order to ascertain the size of segments that will allow for the most reliable and differentiating mean segmental TTR. The above mentioned study by Chotlos (4) is concerned with this problem also.

The cumulative TTR curve. A curve of the cumulative TTR for a given language sample can be plotted by computing successive TTR's as increments are added to the sample. For instance, the cumulative TTR for a 1000-word sample would be plotted as follows: TTR values are to be represented along the ordinate and number of words along the abscissa. The abscissa values may be in units of one word, or ten words, or 100 words, etc., as desired. If the unit is one word, 1000 TTR's would be computed in plotting the cumulative curve for the 1000-word sample; if the unit is ten words, 100 TTR's would be computed; if the unit is 100 words, ten TTR's would be computed, etc. Thus, if the unit is ten words, the first value will represent the TTR for the first ten words of the sample, the second will represent the TTR for the first 20 words, the third will represent the TTR for the first 30 words, etc. The problem of fitting an equation to the resulting curve is dealt with in some detail by Chotlos (4). Basically, the problem concerns the relation between D (number of different words, or types) and N (number of words, tokens) in the given sample. This problem has been given considerable attention by Zipf (20), Carroll (3), and Skinner (16). The relevant data presented by Chotlos (4) indicate the degree to which the relation of D to N promises a means of predicting vocabulary, in the sense that the value of D for a given N provides a basis for predicting D for a specified N of larger magnitude.

The decremental TTR curve. Sup-

pose a 1000-word sample to be divided into ten 100-word segments. The TTR is computed for the first segment. Then, the number of different words in the second segment that did not occur in the first segment—i.e., the number of *new* types introduced in the second segment—is found. The TTR for the second segment is then computed by dividing this number—not the number of types, but the number of *new* types—by 100, which is the number of tokens in the second segment. In the same way, the TTR's for the third, fourth, and each of the other segments may be computed, by dividing the number of tokens, 100 in each case, into the number of *new* types introduced into the sample for the first time in the segments under consideration. The resulting curve of these successive segmental TTR's may be expected to show a relatively steeper slope than the cumulative TTR curve, and the measure representing the slope of this curve may be found to be of special interest. It represents, of course, the rate of decrement in the use of new types, the rate at which the individual "uses up" his vocabulary in producing a language sample. Decremental TTR's should represent in a peculiarly direct quantitative manner one aspect of language development, when applied to language samples secured successively from the same children. The decremental TTR curve is, of course, the first derivative of the cumulative TTR curve, and thus it is not actually necessary to fit a curve to the decremental TTR data if the cumulative TTR curve has been computed.

Type-frequencies. A simple objective language measure is that which expresses the frequency of occurrence of each different word, or type. Such frequencies, as reported for large samples of written language by Thorndike (17), Horn (6), and others, have been used chiefly in the preparation of school readers, spelling books, etc. Certain other uses of such data are obvious. When type-frequencies are based on the kinds of language samples to be used in the present program they may be regarded as representing language behavior norms. In previous studies of word-frequencies it would seem that the primary objective has been simply to determine the relative frequency of occurrence of each word, and with some exceptions special interest has attached to those words which have been found to occur with especially high frequencies. The main objective of the present program in this connection is somewhat different. Chief interest lies in ascertaining individual and group differences in the relative frequency with which particular kinds of words are used. One may determine (a) type-frequency changes that characterize language development; (b) type-frequency characteristics of the language of special groups, especially those that may be found to differentiate one group from another, as schizophrenics from normal subjects, scientists from novelists, etc.; (c) the particular type-frequencies that correlate significantly with such other variables as intelligence, emotional stability, educational level, etc. Attention may be given to the following types of words (and to any others that may be found to be useful):

a. Self-reference words.
b. Quantifying terms (precise numerical words).
c. Pseudo-quantifying terms (words loosely indicative of amount, size, etc., such as *much, many, lots;* or *very, highly,* etc., used as qualifiers of other pseudo-quantifying terms, as in such expressions as "very much").
d. "Allness" terms (superlative or extreme

words, such as *never, always, all, nobody, everyone,* etc.).

e. Words expressive of negative evaluation, such as *no, don't,* etc., and *horrid, unsatisfactory, dislike,* etc.

f. Words expressive of positive evaluation.

g. Qualification terms (words that serve to qualify or limit statements, such as *except, but, however, if,* etc.).

h. Terms indicative of consciousness of abstracting (such words as *apparently, seems, appears, as if, to me,* etc.; as indicated by the last two examples, for purposes of this type of analysis it will be necessary to treat certain phrases as single words. What we call the dogmatic or "closed mind" attitude might be expected to be characterized by language in which these terms are relatively lacking.)

Ratios of any one of the above types of words to any one of the other types might be computed for given language samples, and their significance evaluated. The ratio of the terms indicative of consciousness of abstracting to "allness" words, for example, might be expected to differentiate individuals and groups in ways that should be of theoretical and practical importance in the study of personality.[1]

The relative frequency of use of the various grammatical types of words—nouns, adjectives, verbs, adverbs, etc.—might also be determined, as well as ratios of nouns to adjectives, adjectives to verbs, verbs to adverbs, nouns to verbs, adjectives to adverbs, etc., and the ratio of these four to all other words. With language development, the relative frequency of nouns particularly and also of verbs may be expected to decrease with reference to the relative frequency of adjectives and adverbs. The degree to which these and other possible relationships can be utilized as measures of language development and of individual and group differences should be ascertained. Busemann (2) and Boder (1) have employed the adjective-verb quotient to indicate certain kinds of personality differences, and to differentiate samples of written language. Sanford (13) has reported a personality study involving this and other related measures. The present series of studies involves analyses in this general connection. Mann (11) applies the adjective-verb quotient and also adjective-noun and adverb-verb quotients in her comparative study of the written language of schizophrenic patients and university freshmen. Fairbanks (5) investigates the relative frequencies of occurrence of various parts of speech in comparing the spoken language of schizophrenic patients and university freshmen. Chotlos (4) presents similar data in terms of types and tokens, respectively, and he also presents TTR values for nouns, verbs, adjectives and adverbs, respectively, for written language samples obtained from Iowa school children.

Proportionate vocabulary. How many different words or types make up 25, or 50, or 75 per cent of a given language sample? In the study by Fairbanks (5), 30,000-word samples of spoken language were obtained from schizophrenic patients and "superior" university freshmen, respectively. For the freshmen just 46 different words or types comprised 50 per cent of the 30,000-word sample, and for the schizophrenic patients this figure was 33 types. This is the more striking, perhaps, when expressed by saying that for the schizophrenic patients approximately one-tenth of one per cent of the words made up 50 per cent of the total

[1] Reference made here to "allness" terms and to the notion of consciousness of abstracting implies the writer's debt to Alfred Korzybski. See especially Korzybski, A., *Science and Sanity, An Introduction to Non-Aristotelian Systems and General Semantics,* Lancaster, Pa.: Science Press, second edition, 1941.

sample. In fact, *one* word, the one most frequently used by the schizophrenics, which was the word *I* made up slightly over 8.3 per cent of their entire 30,000 words.

A sample of, say, 1000 words might be analyzed in such a way as to yield a curve as follows: Along the ordinate percentages would be represented; these percentages would correspond to numbers of tokens. For example, suppose that 100 tokens make up 10 per cent of the 1000-word sample; it is this 10 per cent and other percentage values so computed that would be represented along the ordinate. Other percentages would lie along the abscissa; these percentages would correspond to numbers of types. Thus, suppose that 10 types comprise 1 per cent of the total of 1000 tokens; this 1 per cent and other percentage values so computed would be represented along the abscissa. The curve showing the relation between these two sets of percentages would be made up of points expressing such values as the one cited above: for the schizophrenic patients 0.1 per cent of the words (this percentage representing types) made up 50.0 per cent of the sample (this percentage representing tokens). The relation symbolized by this curve can be expressed mathematically, of course, and it is proposed to examine its usefulness as a basis for comparing different language samples or any given sample with a norm or standard sample. The relationship discussed here can be expressed, of course, in terms of rank and frequency. That is, a curve that is fitted to word-frequencies as a function of rank, the most frequent word having the lowest rank number, 1, represents in an alternative way the same phenomenon that is discussed here in terms of proportionate vocabulary. (See Zipf [20].)

Standard frequency vocabulary. The word counts that have been published by previous workers, and the one to be done in the present program, can be used separately or pooled in arriving at a standard frequency-of-use rank number for each different word included in them. Such rank numbers would represent the relative frequency with which each word had been used in the total language sample—presumably drawn from a more or less representative population of individuals—not in terms of the actual number of times each word was used, but in terms of its rank. Thus, the most frequently used word would have a rank number of 1, the next most frequently used word would have a rank number 2, etc.

With the resulting table of rank numbers, it would be possible to score any given language sample by noting the rank number of each word (token) contained in it, and computing the mean (or median) of these rank numbers. The lower the mean of the sample the more heavily loaded it is with words that are used relatively frequently by people generally. We may say, then, that this mean rank number of a language sample represents the "standard frequency vocabulary" employed in it. It is to be reasonably expected that language development would be characterized by increase in this measure, and that the measure would serve to differentiate individuals and groups.

A less refined, and perhaps nearly as adequate, form of this measure could be worked out in terms of standard frequency rank numbers on a categorical basis. That is, the first 100 most frequently used words, for example, could all be given the same rank number, the number 1, the second 100 words could all be assigned the rank number 2, etc.

Statistical analysis may indicate advantages in classes or categories of unequal magnitudes, putting the more frequently used words in smaller groups, for instance, and the less frequently used words in larger groups, or vice versa, perhaps varying the number of words in a group in some relation to the frequency with which they are used. Comparison of results obtained from use of various forms of the measure will determine the relative merits of each.

Verbal output. A very simple language behavior measure is that which expresses the verbal output of an individual. Individual differences and intra-individual variations with respect to verbal output are, of course, obvious. Their significance in relation to the various aspects of personal and social adjustment have not been thoroughly or systematically investigated. It is planned to include an attempt in this direction in the present research program.

Verbal output is not meant to be synonymous with speaking or reading rate, as that term is used to refer to verbal output under relatively optimal conditions. An individual's verbal output under various conditions may, and usually does, fall considerably under what it is when he speaks at or near his optimal steady rate. Verbal output may be expressed, of course, in terms of rate.

The measure may express number of words spoken or written per unit of time, or in response to a specified stimulus under standard conditions. It may also express the proportion of a time unit during which an individual produces spoken or written language. For example, two individuals could be compared by placing them together for one hour and recording (a) the speaking time of each, (b) the total number of words spoken by each, and (c) the verbal output of each in terms of words spoken per minute. It is to be noted that these measures are different from a measure of the *rate* of verbal output *while speaking.* It would be of interest, of course, to correlate such a measure of rate with the other verbal output measures.

Word length. Since the studies of Zipf (20) have shown word length to be highly correlated negatively with frequency of use—the shorter the word the more frequently it occurs—it is not planned at this time that measures of word length will be included to any important degree in the present program. It is mentioned here, however, because the data to be utilized will be so tabulated that word length could be studied if findings indicate that this would be advisable. It is a rigorously objective and highly reliable measure (15).

Sentence length. Sentence length is a measure that presents serious operational difficulties in the study of spoken language, although it may be generally satisfactory in the analysis of written language. It is planned to include it in the analysis of at least a selected set of the written language samples.

SPECIAL TYPES OF LANGUAGE BEHAVIOR TESTS

The *Extensional Agreement Index* (*EAI*) expresses the degree of agreement among *n* persons in defining a given term extensionally—i.e., by pointing to or exhibiting somehow the actual objects, phenomena, etc. to which the term refers.[2] Thus, the kind of behavior which the EAI is designed to measure is not observation so much as word-fact relat-

---

[2] This measure was introduced and briefly discussed in Johnson, W., *Language and Speech Hygiene,* referred to in footnote No. 1.

ing. The EAI may range in numerical value from 0.0 to 1.0, 0.0 representing no agreement and 1.0 representing maximum possible agreement among *n* persons in relating or applying a given word as a label to actualities. Its theoretical and practical significance lies in the fact that it makes possible not only an index of a person's conformity or idiosyncrasy in his extensional use of words, but also a measure of the degree to which any given term may be regarded as testable, or extensional, or operational—or vague. If in the statement "Stutterers are psychoneurotic" the term "psychoneurotic" has an EAI of, say, .18, the statement is not to be regarded as highly testable or factually meaningful, since *n* persons would disagree considerably as to just what is to be observed in order that the validity of the statement might be tested. The EAI offers, therefore, a means of quantifying to some degree such notions as are represented by the terms "verifiable," "operational," etc.

The EAI may be computed in several different ways. Tuthill (18) in a study made as part of the present program demonstrated a variety of ways of computing such a measure of extensional agreement. The basic formula is

$$\text{EAI} = \frac{x}{y}$$

in which x represents the number of obtained agreements and y the maximum possible number of agreements. The EAI, then, represents the per cent of the maximum possible number of agreements that are obtained in a given case.

For example, imagine four different pictures and ten different persons who are each asked to apply the label "most artistic" to one of them. Suppose the label is applied to picture A by 3 persons, to picture B by none, to picture C by 5, and to picture D by 2. If there had been perfect agreement, all 10 persons would have applied the label to the same picture. Thus, the number of agreements among the 10 persons that would have occurred under these conditions is to be regarded as the maximum possible number of agreements. This number may be determined by the formula $(n - 1)\ .5n$ and since $n = 10$, the maximum possible number of agreements is $9 \times 5 = 45$. The number of agreements actually obtained is to be computed as follows: The three persons who applied the label to picture A agreed 3 times, since when $n = 3$, $(n - 1)\ .5n = 3$. There were no agreements with regard to picture B, in terms of the technique for computing the EAI that is here being used. Using the formula $(n - 1)\ .5n$, there were 10 agreements in the labeling of picture C, and one in the labeling of picture D. In all, then, 14 agreements were obtained. Therefore, EAI = 14/45 = .31, which may be interpreted as indicating that the number of agreements obtained was 31 per cent of the maximum possible number.

This is an example of an extremely simple case, used to illustrate the application of the basic formula. Another example will serve to indicate an important modification of the basic formula. On July 9, 1939, the American Institute of Public Opinion released to newspapers the results of a survey in which each of several thousand persons had been asked to apply one of the labels, "Conservative," "Liberal," and "Radical," to each of ten prominent Americans.[3] The results were presented

[3] The material upon which the present discus-

in percentages as follows:

	Conservative %	*Liberal* %	*Radical* %
Hopkins	4	55	41
Roosevelt	1	62	37
La Guardia	8	64	28
Farley	13	63	24
Dewey	45	47	8
Hull	51	46	3
Garner	64	32	4
Vandenberg	67	29	4
Taft	86	13	1
Hoover	92	5	3

These figures represent only the labeling reactions of persons "who knew or had some idea of the terms when later in the survey they were asked pointblank what the words . . . meant." From these data it is possible to compute an EAI for each of the three terms involved. The procedure to be used will differ in three important respects from that used in the above example of the four pictures. In the first place, in the first example there was only one label to be applied by each of ten persons to only one of four possible referents. In the present case, there were three labels, any one of which was to be applied to each of ten referents. In the second place, there were ten labelers in the first example; in this one there were many thousands, and the numbers have been converted into percentages. These percentages will be used instead of the raw numbers in computing the EAI's. In $(n - 1)\ .5n$, n will represent 100 in computing the maximum possible number of agreements. Lastly, instead of assuming, as was done in the first example, that agreements occur only when labels are applied, and not when they are not applied, we shall assume that both the application of a label and the refusal to apply it may involve agreement. When this assumption is made, the *net* number of agreements involved in the application and non-application of a given label to a given referent can be computed as follows. Let $x =$ the number who apply the label, and $n - x$ the number who do not apply it. Then, the number of agreements among those who do apply the label is found by the formula, $(x - 1)\ .5x$. Similarly, the number of agreements among those who do not apply the label equals $(n - x - 1)\ .5\ (n - x)$. The *net* number of agreements is found simply by subtracting the smaller of these values from the larger. And the EAI is found by dividing this *net* number of agreements by the maximum possible number of agreements. Thus,

$$\text{EAI} = \frac{\dfrac{(n - x - 1)\ (n - x)}{2} - \dfrac{(x - 1)\ x}{2}}{\dfrac{(n - 1)\ n}{2}} = \left|\frac{2x - n}{n}\right|$$

In this way, the EAI of each given term is computed for each referent, and the EAI's of the term for the various referents (in the present case, 10) are averaged. For the term "Liberal" the following results were obtained:

"Liberal"

	% Labeling	$\left\lvert\dfrac{2x - n}{n}\right\rvert$
Hopkins	55	.10
Roosevelt	62	.24
La Guardia	64	.28
Farley	63	.26
Dewey	47	.06
Hull	46	.08
Garner	32	.36
Vandenberg	29	.42
Taft	13	.74
Hoover	5	.90
Ave. EAI		.34

The obtained number of agreements was, on the average, only 34 per cent of the number representing complete agreement as to the extensional meaning of the word "Liberal," as applied or not, to the ten men listed, by the presumably

sion is based appeared under the copyright of the American Institute of Public Opinion in the *Des Moines Register*, July 9, 1939.

random sample of persons surveyed by the Gallup organization in the summer of 1939. The variability is of interest. As applied to Hopkins, Dewey and Hull, the term "Liberal" proved to be almost entirely meaningless; there was virtually no agreement as to whether these men were or were not suitable referents of the term. There was relatively high agreement, on the other hand, that Taft and Hoover were not to be labeled "Liberal." The mean EAI was .58 for "Conservative" and .69 for "Radical."

Dr. Gallup, under whose name the survey report appeared in the press, did not, of course, report his findings in terms of these EAI's. Moreover, he seems to have missed a basic point, in stating that the survey results indicated "the way American voters—*rightly* or *wrongly*—are classifying the figures in United States political life." (The italics are the present writer's.) The words "rightly or wrongly" seem to imply the assumption that there is a "right" way and a "wrong" way to apply such a label as "Liberal," that such a term has somehow an intrinsic "meaning," presumably known by some means to someone somewhere, quite aside from and more valid than the extensional meanings ascribed to it by those persons who actively relate it to various referents. There would appear to be, from an extensional point of view at least, no "right" or "wrong" about it, except in the sense that in matters of this kind one might (or might not) prefer to assume that the majority is "right." Be that as it may, however, Dr. Gallup carried out, in this particular survey, what amounted to a very ambitious effort to determine by vote the extensional meanings of a group of words. And by using his results to compute the EAI's of these words, it becomes possible to measure fairly precisely the vagueness or factual meaningfulness of some of our important political terms.

The resulting EAI's afford a degree of insight into the processes of political controversy, and point to one of the fundamental problems in connection with social organization. The EAI of .34 for "Liberal" strongly suggests that such a statement as, "America should (or should not) have a liberal in the White House," is to be regarded as essentially "lyrical." Like our remarks about the weather, which are not to be mistaken for meteorological reports, the remark that "So-and-so is a liberal" is not to be regarded as a statement chiefly descriptive of So-and-so. For the most part, it merely serves to announce one of the ways in which the speaker proposes to apply the word "Liberal," and thus it is mainly indicative of an aspect of the speaker's language behavior. To know the EAI of a word, as computed from data as adequate as those provided by Dr. Gallup, is to know something quite precise and significant about the language behavior of a speaker or writer who uses it, particularly if he gives no indication of awareness of the word's descriptive limitations, as these are implied by its EAI. The descriptive limitations of a word with an EAI of .34 are probably so great as to render it practically meaningless referentially in many contexts. It is to be regarded as being in many instances little more than noise or ink marks, meaningful chiefly in being symptomatic of the speaker's or writer's neurosemantic state. That is to say, it is more revealing as behavior than as language; it symbolizes the speaker more than it symbolizes anything he may appear to be speaking about.

This rather long discussion of the EAI has been given in order to make more or less clear, not only the basic operations

involved in its computation, but also certain of its implications. The EAI of a term, computed from data obtained under adequate conditions, is indicative of one of the most important characteristics of word usage, the relatively precise degree to which words may be regarded as factually meaningful—or vague.

Use of the measure requires that it be computed from data obtained under known and specified conditions; moreover, the particular form of the basic formula to be used in computing it will vary somewhat with the nature and purpose of the investigation. It is proposed that preliminary work to be done in the present program will involve construction of a test by means of which EAI's for a number of different terms can be determined under a variety of conditions. Work already done indicates that the reliability of such a test can be expected to be quite high, that its administration and scoring offer no insurmountable problems, and that data obtained by means of it will reveal differences between words and between individuals and groups.

In the administration of this test it is planned that the subject will be given a word in a standard context, as in the statement: "Point to the pictures that show people doing *good* things." The subject then points to such pictures, among a standard set of pictures, as to him represent referents of *good* as so used. Each picture in the set is numbered and the number of each picture to which the subject points is recorded. From data so obtained from each of a group of subjects, the EAI of each word in the test is to be computed, as was done for the Gallup poll data presented in the preceding pages.

As part of the present program, a study has been made by J. Wilson and the writer (7) in which graduate students and instructors in psychology defined extensionally, by reference to a list of statements taken from psychology texts, the terms "law," "theory," and "hypothesis." The mean EAI's obtained were .62 for "law," .40 for "theory," and .28 for "hypothesis."

Extensional Synonymity Index (*ESI*). Such EAI's represent the relative degree of vagueness of words as used. By treating the test data in other ways, they can be made to yield two other types of information, represented by an extensional synonymity index (ESI) and an extensional conformity index (ECI), respectively. By recording the percentage of all the subjects who point to each picture, or other types of referent, in defining each word, it is possible to measure the degree of synonymity between any two words. The formula $\mathrm{ESI} = \frac{c}{\sqrt{xy}}$, in which c represents the percentage of subjects pointing to a given picture in defining both of two given words, and x and y represent the percentages of subjects pointing to the picture in defining each of the two words, respectively. This value is to be computed for each picture, and the values thus obtained for all the pictures are to be averaged in deriving an expression of the mean degree of synonymity between any given pair of words.

Extensional Conformity Index (*ECI*). The percentages of subjects pointing to each picture in defining each word can also be used as word-fact relating behavior norms. Thus, the pictures may be "weighted" according to these values, and on the basis of them the pointing or labeling of a given individual can be evaluated. For example, if a given individual in defining the word "good" were to point to certain pictures, he

would be showing less conformity to the group that he would be in pointing to certain other pictures. The mean of the percentage values of the pictures to which an individual points in defining a given word would represent his degree of conformity to the group in his extensional use of that word. We may call this his extensional conformity index (ECI), and individual differences expressed in terms of the ECI might be found to be a factor in personality adjustment.

The Intensional Agreement Index (IAI) expresses the degree of agreement among *n* persons in defining a given term intensionally—i.e., by giving its verbal equivalents. A dictionary definition is to be regarded as an intensional definition, as the term is here used. Like the EAI, the IAI may range in value from 0.0, representing no agreement, to 1.0, representing maximum possible agreement.

In a preliminary study carried out by N. Whitman and the writer (8), an attempt was made to determine IAI's for each of certain terms used in the field of psychology (*learning, perception, emotion,* and *personality*) and certain terms used in the field of biochemistry (*fats, lipids, enzymes, oxidation,* and *basal metabolism*). Textbooks in each field were examined until for each term six definitions (from six different authors) had been found. These definitions were then edited so as to exclude all words except nouns, verbs, adjectives, and adverbs (the adverbs *when* and *where,* the adjectives *that, these, those,* and *which,* and articles used as adjectives were also excluded). Then for each term the number of types (different words) used in all six definitions was recorded, and the number of definitions in which each type was used was determined. The number of obtained agreements, in the use of any given type by the six textbook authors, was found by means of the formula (n − 1) .5n, in which n represents the number of definitions in which the type occurred. The values thus obtained for the various types were summed in determining the total number of obtained agreements shown by the six textbook writers in verbally defining the term in question. The maximum possible number of agreements was computed by using the formula x(n − 1) .5n, in which n represents the total number of definitions, six in each case, and x represents the total number of types used in all the definitions. The maximum possible number of agreements was then divided into the obtained number of agreements in determining the IAI of a given term. The IAI's as thus determined, were:

		No. of types
Psychological terms		
Learning	.024	44
Perception	.006	40
Emotion	.010	48
Personality	.007	46
Ave.	.012	44.5
Biochemical terms		
Fats	.080	56
Lipids	.150	59
Enzymes	.127	20
Oxidation	.067	27
Basal Metabolism	.035	50
Ave.	.092	42.4

The difference between the mean IAI's may be regarded as indicating a measurable difference between the fields of psychology and biochemistry with regard to the degree of terminological agreement that has been achieved in them to date. One important aspect of scientific development is to be observed in the fact that among biochemists at the present time there is a tendency to abandon the term *fats* in favor of the term *lipids*—a tendency to replace one term with another that has a higher IAI. Increasing agreement as to definitions, both intensional and extensional,

is a basic characteristic of the development of a science; and a means of measuring the degree of agreement that has been achieved within the various fields makes possible a peculiarly objective comparison of them in this important respect. Degree of similarity among verbal formulations generally can be measured in terms of the IAI.

The procedure followed in the above study of psychological and biochemical terms can be modified in at least three ways. First, the definitions can be obtained directly from the subjects rather than from text books or other published material. Second, the subjects can be instructed to define each word by listing synonyms of it, and the number of synonyms to be listed can be limited. Third, the words to be defined need not be presented only in isolation, but they may be presented also in context, other words to be substituted by the subject for the word in question, or a definition to be written for the word as used in the particular context. The influence of differences in context on the meaning, and on agreement as to the meaning, of specific words can thus be investigated.

Intensional Synonymity Index (*ISI*). From data of the type just discussed it is possible to obtain measures of intensional synonymity. Degree of synonymity of given pairs of words defined extensionally can be measured by means of procedures already described. Similar procedures can be used in the present connection. For example, suppose the words *good* and *worthwhile* to have been defined by each of 100 subjects, each of whom defined each word by listing three synonyms for it. The degree of intensional, or verbal, synonymity between these two words can then be computed by means of the formula $\frac{c}{\sqrt{xy}}$ in which c represents the number of terms (types) given by the 100 subjects as synonyms for both words, and x and y represent the number of terms (types) listed as synonyms for each of the two words, respectively. The correlation between extensional and intensional synonymity indexes would be of interest.

Semantic vocabulary test. As has been indicated previously in this outline, vocabulary measures can be obtained from a language sample obtained from any given individual in terms of type-token ratios, type frequencies, proportionate vocabulary, and standard frequency vocabulary. Another type of vocabulary test might be attempted. A common criticism of ordinary vocabulary tests is that while they are indicative of the number of words an individual "knows" or "recognizes," they are not necessarily indicative of the range of "depth" of the individual's knowledge of or skill in using each word that he "knows." The problem raised by this criticism involves technical difficulties, but certain approaches to its solution appear to be possible (14).

Investigation could be made of the feasibility of constructing a vocabulary test of such a nature that the individual's ability to use each word would be sampled in detail. It is possible to distinguish types of meaning, such as meaning in terms of use, variety, differentiating characteristics, sources, etc. For example, the word *orange* can be defined in terms of (a) the various uses of oranges, (b) the kinds of oranges, (c) the characteristics that differentiate oranges from other things, (d) the geographical areas where oranges are grown, the methods by which they are grown, the history of these methods, etc., and (e) in terms of the scientific research that has been done on oranges, the methods used in

picking, packing, processing, marketing, transporting, etc. This does not exhaust the problem of defining *orange,* but it illustrates the possibility of devising a vocabulary test of a type that should make possible a measure of vocabulary "depth" as well as "range."

Measures of "allness." Previous mention has been made of "allness" terms, such as *all, everyone, nobody, every, never, absolutely,* etc. Language spoken during moments of anger or despair, or other relatively profound affective states, appears to be particularly characterized by such terms. They give to language a character which reflects what is usually referred to as dogmatism, or stubbornness, inflexibility, etc. Orientation on the basis of dichotomies, or of the excluded middle—a two-valued, either-or orientation—appears to be basic to and to be fostered by, this sort of language. The degree to which one is prone to two-valued orientation is probably an important aspect of one's general adjustment, personality development, intelligence, etc. Insofar as it might prove possible to set up rigorous criteria of allness terms, the frequency of their use in language samples could be studied.

Another approach to the study of allness, however, is also to be proposed. From one point of view allness may be regarded as manifested in extreme responses in situations where they are not mandatory. An attempt could be made to construct a reliable test involving, say, 100 items, to each of which a response can be made along a graduated scale expressive of extreme and intermediate degrees of preference, attitude, behaviorial tendency, etc. At least five and possibly seven or more alternative responses to each item should be provided, one expressive of neutrality or average tendency and the others distributed on either side and graduated toward the two extremes. The test would be scored, not in terms of the preferences, etc. expressed, but in terms of the proportion of extreme (allness) responses. It is anticipated that two main types of evaluative tendencies might be indicated by such a test, the tendency to give extreme responses, or allness, and an extreme tendency to give indecisive, indefinite, neutral responses. The latter might characterize certain schizoid conditions, for example. It is to be noted that this type of test should get away from one common weakness of pencil-and-paper tests, in that the effect of falsified responses on the score will be minimized, since the "intensity" rather than the "content" of the responses will determine the score.[4]

Tests of verbal differentiation. It would appear reasonable to assume that the adequacy of generalization or "abstract thinking" depends largely upon the adequacy of the analysis or differentiation upon which the generalizing is based. This is indicated by an examination of practically any generalization process; it is especially obvious, perhaps, in medical diagnosis. The ability to observe, respond to, and relate differences would appear to limit the ability to abstract similarities effectively. In fact, abstracting (roughly, generalizing) can be defined as a process of leaving out details or differences; similarities are recognized and formulated in accordance with the way differences are disregarded, not observed, or related. Consciousness of abstracting (9, 10), therefore, in any given instance, is seen to depend on an awareness of the differences that are being disregarded or related in the abstracting of similarities.

[4] Previous work suggestive of this approach has been reported by Watson (19).

It is proposed to construct a test specifically designed to measure an individual's ability to express differences, or to perform verbal differentiation. It is the intention to begin with the simple procedural plan of presenting the subject with pairs of objects, designs, etc. and requesting him to tell the differences between them. A time limit, to be determined, is to be set for each response. An attempt is to be made to score the responses in each of three ways. First, the mere length of response is to be measured; it is hardly to be expected that this will suffice, except possibly as a very gross measure. Second, the number of differences enumerated is to be noted; it will be necessary to formulate rigorous criteria of a "difference." Third, various forms of the type-token ratio are to be tried as possible expressions of the subject's level of performance.

Assuming the construction of a reliable test, scores on the test are to be related to other variables. The relation of differentiating ability to intelligence, as measured by current standard tests, and to other criteria of competence, is of particular interest.

SUPPLEMENTARY MEASURES

The entire research program here proposed involves not only the language behavior measures discussed above, but also certain other measures which are to be used in order to obtain data concerning the relation of the language measures to other aspects of behavior. Among these supplementary measures are tests of intelligence, measures of mental and chronological age, achievement and aptitude tests, measures of silent and oral reading, of speech and writing, and various indices of personality.

STUDIES COMPLETED

To date six studies have been completed, and a considerable amount of preliminary or exploratory investigation has been done. The six completed studies have been done by Fairbanks (5), Mann (11), Chotlos (4), Tuthill (18), Johnson and Whitman (8), and Johnson and Wilson (7). The investigations so far completed have been concerned mainly with problems of method, although they have been designed to contribute, also, to a fuller understanding of language behavior in its various relationships.

Individual 3,000-word spoken language samples were obtained by Fairbanks (5) from each of 10 schizophrenic patients and 10 university freshmen. Mann (11) obtained 2,800-word written language samples from each of 24 schizophrenic patients and 24 university freshmen. The writer has obtained 3,000-word written samples from each of approximately 1,000 Iowa public school children, selected on the basis of age, sex, I.Q., type of school (rural, town, city) and socio-economic status.[5] A selected set of 108 of these written language samples have been analyzed in considerable detail by Chotlos (4).

The studies which follow the present article in this monograph will serve to illustrate some of the above types of approach to the investigation of language behavior. The present discussion is offered as a general introduction to these and to the further studies that will, it is hoped, be included in the program of research which has here been outlined.

[5] Acknowledgement is hereby made of grants from the Work Projects Administration for Iowa Projects 4,892 and 5,960, by means of which these data were obtained and tabulated, and to Professor George D. Stoddard, then Director of the Iowa Child Welfare Research Station, who sponsored the projects.

BIBLIOGRAPHY

1. BODER, D. P. The adjective-verb-quotient: a contribution to the psychology of language. *Psychol, Rec.*, 1940, *3*, 310-343.
2. BUSEMANN, A. *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik*. Jena: Fischer, 1925, pp. x + 98.
3. CARROLL, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.*, 1938, *2*, 379-386.
4. CHOTLOS, J. W. Studies in language behavior: IV. A statistical and comparative analysis of individual written language samples. *Psychol. Monogr.*, 1944, *56*, 75-111.
5. FAIRBANKS, H. Studies in language behavior: II. The quantitative differentiation of samples of spoken language *Psychol. Monogr.*, 1944, 56.
6. HORN, E. *A Basic Writing Vocabulary: 10,000 Words Most Commonly Used in Writing*. Univ. of Iowa Monogs. in Educ., 1926, No. 4, 225 pp.
7. JOHNSON, W., & WILSON, J. The degree of extensional agreement among twenty psychologists in their use of the labels *hypothesis, theory* and *law*. Unpublished research.
8. JOHNSON, W., & WHITMAN, N. Intensional agreement indexes of certain terms in psychology and biochemistry. Unpublished research.
9. JOHNSON, W. Language and speech hygiene. *Gen. Semantics Monogr.*, No. 1, Chicago: Institute of General Semantics, second edition, 1941.
10. KORZYBSKI, A. *Science and Sanity, An Introduction to Non-Aristotelian Systems and General Semantics*. Lancaster, Pa.: Science Press, (2nd ed.), 1941.
11. MANN, M. B. Studies in language behavior: III. The quantitative differentiation of samples of written language. *Psychol. Monogr.*, 1944, *56*, 41-74.
12. SANFORD, F. H. Speech and personality. *Psychol. Bull.*, 1942, *39*, 811-845.
13. SANFORD, F. H. Speech and personality: a comparative case study. *Character & Pers.*, 1942, *19*, 169-198.
14. SEASHORE, R. H., & ECKERSON, L. D. The measurement of individual differences in general English vocabularies. *J. Educ. Psychol.*, 1940, *31*, 14-38.
15. SKALBECK, O. M. A statistical analysis of three measures of word length. M.A. Thesis, Univ. Iowa, 1938.
16. SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, *1*, 71-76.
17. THORNDIKE, E. L. *The Teacher's Word Book*. New York: Teachers College, Columbia, 1921, 134 + vi pp.
18. TUTHILL, C. E. A quantitative study of extensional meaning with special reference to *stuttering*. Ph.D. Thesis, Univ. Iowa; 1939.
19. WATSON, G. B. *The Measurement of Fairmindedness*. New York: Teachers College, Columbia Univ., 1925, pp. 97.
20. ZIPF, G. K. *The Psychobiology of Language*. Boston: Houghton Mifflin Co., 1935, pp. ix + 336.

II. THE QUANTITATIVE DIFFERENTIATION OF SAMPLES OF SPOKEN LANGUAGE

II. The Quantitative Differentiation of Samples of Spoken Language[1]

(Cornelius)

HELEN FAIRBANKS
University of Iowa

I. INTRODUCTION

The general program of research, of which this study is a part, has been outlined in a previous article by Johnson (9). As the first study to be undertaken within that program, this investigation is concerned primarily with problems of method. It is concerned specifically with a partial exploration of the possibilities of measuring certain aspects of language behavior, and of differentiating samples of spoken language in terms of the measures employed.[2] It is assumed that the first step in a comprehensive language behavior research program lies in the attempt to develop adequate techniques of measurement necessary for the formulation and testing of hypotheses.

In accordance with this point of view, an attempt was made in this study to obtain two groups of language samples which might be assumed to be sufficiently different as to make possible some indication of the sensitivity of the measures to be employed. On the basis of this consideration it was decided to obtain the language samples from hospital patients suffering from schizophrenia, and from university freshmen who were judged to be superior in terms of criteria to be indicated in the section on Procedure. The 'superior' freshmen were chosen with the expectation that they would furnish relatively 'adequate' language, and the schizophrenic patients were used on the assumption that their language would prove to be relatively 'inadequate', and that the contrast might be sufficiently marked to be quantitatively expressed. One of the important clinical manifestations of schizophrenia is to be noted in the language of the patients suffering from the disease (13). As the illness progresses there is a tendency for the language to appear disconnected, illogical, even incomprehensible. Stereotypy in verbal expression is frequently apparent. Thus, there would seem to be reasonable ground for expecting that the language of schizophrenic patients might be demonstrably different, quantitatively, from that of 'superior' normal subjects. Relevant studies have been reported by White (15), Woods (16), and Cameron (2, 3).

It was an incidental consideration that any differences that might be revealed as between these two groups would possibly be of psychiatric and psychological interest. The primary purpose of the investigation is, however, methodological, and any conclusions to be drawn from the findings with regard to the nature of 'schizophrenic language' are to be most carefully evaluated. In this

[1] This study was done in the Department of Phychology at the State University of Iowa as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy. It is part of a program of research on language behavior and was directed by Wendell Johnson. The writer is grateful to Dr. Andrew H. Woods, Director, and the staff of the Iowa State Psychopathic Hospital [1938-40]; and to Dr. Leonard P. Ristine, Superintendent, and the staff of the Mt. Pleasant State Hospital, for their cooperation in securing subjects for the investigation. Special acknowledgment is made of the assistance of Dr. Frank Robinson, resident psychiatrist at the Iowa State Psychopathic Hospital during 1938-39.

[2] A companion study by Mann [11] is concerned with written language. The specific measures used in that study and this one, as well as several other types of language measures, are discussed in the above mentioned article by Johnson [9].

connection, it is to be emphasized that it was regarded as of first importance to obtain two groups of subjects who might be expected with some assurance to produce demonstrably different language samples. It was for this reason that 'superior' university freshmen were selected. This meant, however, that the possibility of securing schizophrenic patients matched with the freshmen in terms of intelligence and educational background —a difficult undertaking in any case— was deliberately jeopardized. Most of the patients were judged to be of average intelligence or above, as will be indicated later, but the fact remains that any demonstrated language differences between the patients and the students may be due, in part, to differences in intelligence or in scholastic background, and not entirely to schizophrenia, *per se*. It appeared advisable, nevertheless, to establish first, as far as possible, the degree to which the measures used were sensitive or differentiating. Had well matched groups been used and no differences in language found, the basic question of the differentiating value of the measures would have remained unanswered. It could not have been concluded whether there were no differences to be measured, or that the measures used were too crude to reveal them. Therefore, the methodological problem was placed first in importance in designing the study, but schizophrenic patients were used in the hope that, if the measures turned out to be differentiating, some findings of psychiatric and psychological significance might be gained.

II. PROCEDURE

Two groups of adults served as subjects in this study: (1) ten psychotic patients diagnosed as schizophrenic; (2) ten freshmen at the State University of Iowa. The major characteristics of these groups are summarized below.

Of the schizophrenic subjects, four were patients at the Iowa State Psychopathic Hospital, Iowa City, and the other six, three of whom had previously been in the Iowa State Psychopathic Hospital, were committed patients at the Mt. Pleasant State Hospital at Mt. Pleasant, Iowa. These ten patients were chosen on the basis of the certainty of the diagnosis made of them by the psychiatrists and the possibility of securing their co-operation in the proposed interview situation. Data concerning the individual cases are as follows:

Case 1. Diagnosis: schizophrenia, paranoid type. A single male, aged 46 years, 6 months; educated through 9th grade and one year business college; first psychotic episode in 1916, confined to the Mt. Pleasant State Hospital continuously since 1934; scored Intelligence Quotient of 114 on Wechsler-Bellevue Adult Test and 104 on Revised Stanford-Binet Test, Form L, passing the vocabulary test on the latter at the Superior Adult III level; patient inclined to give up easily on tests; psychometrist commented: "intellectual development has been superior."

Case 2. Diagnosis: schizophrenia, hebephrenic type. A single male, aged 31 years; educated through 8th grade; first mental symptoms in 1935, confined in Mt. Pleasant State Hospital since; scored Intelligence Quotient of 24 on Revised Stanford-Binet Test, Form L, but so deteriorated psychometrist felt no estimate of original intellectual level possible.

Case 3. Diagnosis: schizophrenia, catatonic type. A single male, aged 20 years, 8 months; educated through 11th grade; first mental symptoms in 1939, discharged from Iowa State Psychopathic Hospital several months later as much

improved; scored Intelligence Quotient of 62 on Stanford Revision of Binet-Simon Test but excited and distractible; six weeks later when more co-operative scored Intelligence Quotient of 99 on retest and 115 on Revised Stanford-Binet Test, Form L, passing the vocabulary test on the latter at the Superior Adult II level; original level judged by psychometrist to have been "high average to superior."

Case 4. Diagnosis: schizophrenia, paranoid type. A widower, aged 41 years, 9 months; educated through 8th grade; first mental symptoms in 1934, confined in Mt. Pleasant State Hospital since 1938; had an Intelligence Quotient of 76 on Revised Stanford-Binet Test, Form L, scoring slightly below average on vocabulary test; required considerable urging before trying tests; original level estimated by psychometrist to have been "slighty below average."

Case 5. Diagnosis: schizophrenia, paranoid type. A single male, aged 31 years, four months; educated through high school and business college; expressed paranoid ideas in 1930 and 1935 and developed acute symptoms in 1939, discharged after several months from Iowa State Psychopathic Hospital as improved; scored Intelligence Quotient of 87 on Stanford Revision of Binet-Simon Test, passing vocabulary test at high average level; passed vocabulary test on Revised Stanford-Binet, Form L, at Superior Adult I level; original level felt by psychometrist to have been "average intelligence or above."

Case 6. Diagnosis: schizophrenia, hebephrenic type. A single female, aged 36 years, 11 months; educated through high school and two years' college; first mental symptoms in 1932, hospitalized at Iowa State Psychopathic Hospital in 1933, then committed to Mt. Pleasant State Hospital and there since; scored Intelligence Quotient of 83 on Revised Stanford-Binet, Form L, passing vocabulary test at Superior Adult II level; judged by psychometrist to have been originally "at least high average."

Case 7. Diagnosis: schizophrenia, paranoid type. A single female, aged 27 years, 1 month; educated through two years' college; first mental symptoms in 1930, present episode began in 1938, hospitalized at Iowa State Psychopathic Hospital in 1939, then committed to Mt. Pleasant State Hospital and there since; scored Intelligence Quotient of 118 on Stanford Revision of Binet-Simon Test, passing vocabulary test at very superior level, and 138 on Revised Stanford-Binet, Form L, passing vocabulary test at Superior Adult III level; psychometrist commented that intellectual level was "very superior."

Case 8. Diagnosis: schizophrenia, unclassified type. A single female, aged 37 years, 1 month; educated through preparatory school and four years' college, in Biblical seminary at time of first mental symptoms in 1934; hospitalized at Iowa State Psychopathic Hospital, 1939, discharged home after several months as much improved but symptoms gradually returning.

Case 9. Diagnosis: schizophrenia, paranoid type. A married female, aged 45 years, 3 months; educated through 8th grade; first mental symptoms in 1937, hospitalized at Iowa State Psychopathic Hospital, 1939, and discharged several months later as unimproved with advice to commit patient to a state hospital.

Case 10. Diagnosis: schizophrenia, paranoid type. A married female, aged 31 years, 3 months; educated through high school and business school; first mental symptoms in 1939, hospitalized at Iowa State Psychopathic Hospital for several

weeks, then transferred to Independence State Hospital, Independence, Iowa, and there since; scored Intelligence Quotient of 92 on Stanford Revision of Simon-Binet Test, passing vocabulary test at superior level; psychometrist commented: "vocabulary and the quality of her responses indicate superior intelligence".

In summary, the schizophrenic subjects consisted of five males and five females, ranging in age from 20 years, 8 months, to 46 years, 6 months; six had been diagnosed as paranoid, two as hebephrenic, one as catatonic, and one had not been classified. The length of the illness ranged from an acute episode lasting about a month to an illness that began in 1916 and has gradually shown exacerbations since. The educational backgrounds ranged from 8th grade to college graduation; one patient was felt by the psychometrist to be of very superior intelligence, two superior, two high average to superior, one average or above, one slightly below average, and one too deteriorated to permit evaluation of original level. It was not possible to obtain psychometric ratings on the remaining two patients; of these, one graduated from college and one had no training beyond the 8th grade but was considered an excellent business manager by a local attorney.

The freshman students who formed the second group were chosen on the basis of their September, 1938, scores on the Iowa Qualifying and Placement Examinations. All ranked from the 91st to the 99th percentile on Silent Reading, Comprehension, and from the 95th to the 99th percentile in English Training. It can be assumed that the intellectual level of these subjects is probably superior, as a recent unpublished study by Mitchell (12) indicated a correlation of .76 between the Intelligence Quotients of 66 freshmen as scored on the Revised Stanford-Binet, Form L, and their Composite Score on the Iowa Qualifying and Placement Examination, the average Intelligence Quotient being 122. The group of ten freshmen was chosen on the assumption that its members would represent relatively adequate language usage. Six of the freshmen were female and four were male; the age range was from 17 years, 5 months, to 19 years, 1 month. They came from homes in which the following occupations were represented by the wage earners: bank receiver, jeweler, theatre owner, coal miner, postmaster, county superintendent of schools, life insurance agent, lumberman, odd jobs and trucking, industrial engineer and sales manager.

A consideration of the methods to be used in treating the data and of the issues with which the study was involved seemed to indicate that a 3,000-word spoken language sample from each subject would be adequate. In formulating the procedure care was taken to secure samples that would be comparable from subject to subject and group to group. Because of the frequent difficulty found in getting schizophrenic patients to talk readily, the following interview situation was prepared, utilizing 14 proverbs whose efficacy as stimuli has been demonstrated in previous studies done at the Iowa State Psychopathic Hospital. The following instructions were given to each subject:

"I want you to talk about some proverbs today. You know what a proverb is. A proverb is a sentence that teaches a lesson. I am going to read some proverbs to you, and I want you to tell me what they mean. I also want you to describe a situation in which each proverb would apply. For example, the proverb 'Let

sleeping dogs lie' means that we should avoid stirring up old troubles or quarrels. An example of a situation in which this proverb would apply would be, for instance, if you and a friend had quarreled over something several months ago, you should forget it and be friends with him again instead of continuing to quarrel with him each time that you see him. Do you understand what I mean? Now you tell me what this proverb means, 'The early bird catches the worm.'

"Now give me an example illustrating that."

This procedure was continued with each of the following proverbs:

"He who laughs last laughs best."

"A chain is as strong as its weakest link."

"The devil finds work for idle hands."

"Tell me the company you keep and I'll tell you what you are."

"Deeds are males and words are females."

"Like father, like son."

"What you sow you will reap."

"Barking dogs never bite."

"You can't touch pitch without being tarred."

"A crow is known by the company he keeps."

"A fair face may hide a foul heart."

"A prophet is without honor in his own country."

"It is always darkest just before the dawn."

The subjects were asked to continue talking about anything that they wished to after finishing the proverbs. It was difficult to keep the interview situation as simple for the schizophrenics as for the freshmen, as would be expected with psychotic individuals who show so little response to their environment, and it was necessary to stimulate them more frequently with such questions as why they were in the hospital and what they were doing, in order to get the requisite 3,000 words from each. Two patients had to be interviewed a second time in order to get enough words, the second interview continuing where the first had left off. In one of these cases the total number of words still did not approximate 3,000, and as the patient was removed from the hospital by relatives before a third interview could be arranged, his language sample consists of only 2,800 words. The patients were interviewed by a resident psychiatrist at the Iowa State Psychopathic Hospital, while the experimenter interviewed the freshman subjects. All interviews were completely recorded by means of an electrical dictaphone apparatus, consisting of a microphone, amplifier, and two dictaphones. All recordings were continuous. As the microphone was concealed among books and papers on the interview desk, the subjects were not aware of the fact that their speech was being recorded except in the case of one freshman who happened to uncover the microphone. However, it was the opinion of the interviewer that even in this case speech was not disturbed.

The dictaphone records were then transcribed by the experimenter, following the conventional forms of word division and spelling as closely as possible. The neologisms or coined words occasionally introduced by the schizophrenics were spelled as they sounded phonetically. As would be expected, the intelligibility of the records varied in accordance with the amount of intensity and the clearness of articulation used by the various subjects. Each record was played over until the experimenter was reasonably sure of the transcription. All words and sections which were doubtful were omitted.

A study by Betts (1) has indicated that fewer than one per cent of the words of normal speakers recorded by the electrical dictaphone technique are unintelligible. However, the percentage of such words is probably higher in the present study due to occasional mumbling by the patients, but it cannot be stated definitely just how much. As stated before, the experimenter played the records over until reasonably certain of the transcription, omitting all words or phrases that were doubtful.

The language sample of each subject was divided into 30 consecutive segments consisting of 100 words each. A word count was then made for each protocol by placing a tally mark for each different word on tabulation sheets so organized that each 100-word segment could be tabulated individually. The part of speech for each word was designated as it was tabulated. The following rules were followed in determining what constituted a word:

1. Each group of letters separated by spaces on both sides from adjacent groups of letters was counted as a word, even though it might be part of a place name, as in *Des Moines* (two words), an initial, as in *John D. Rockefeller, Jr.* (four words), and abbreviation of a word previously used, as *coop.* for *cooperative,* a spelling of a word previously pronounced, as *p-a-r-d* for *pard* (one word), or a neologism coined by a schizophrenic patient, as *tombody.*

2. Random letters given consecutively by schizophrenic patients, such as *d-t, were* considered as spellings and counted as one word.

3. Any number was counted as one word; for example, *125* was tabulated as one word.

4. A hyphenated word was counted as one word, Webster's New International Unabridged Dictionary (14) being used as the authority as to whether or not a word should be hyphenated.

5. Sounds like *uh* and *er* uttered by subjects during pauses were not considered as words. However, in one case *uh* and *er* were cited by a subject as examples, in which instance they were regarded as words. The sounds *huh, uh huh,* and *hunh uh* were also regarded as words, being tabulated under *what, yes,* and *no* respectively.

6. Each time a word was used as a different part of speech it was counted as a different word. For example, *mine* as a noun and *mine* as a pronoun were tabulated as two different words.

7. Different tenses of a verb having identical spellings were counted as different words. For example, *read,* present tense, and *read,* past tense, were tabulated as two different words.

8. Common nouns and proper nouns having identical spellings were thrown together. For example, the two words, *Death Valley,* were tabulated under the common nouns, *death* and *valley.*

The data taken from these tabulation sheets were organized into three different sections of results: (1) the type-token ratios, (2) grammatical analysis, (3) word frequencies (8, 9).

III. RESULTS

1. *Type-token ratio.* This measure is computed by dividing the number of different words (types) by the total number of running words (tokens). Since the number of different words decreases as successive increments are added to a language sample (4), the number of tokens used in computing the type-token ratio must be kept constant in order to determine any variations within any given language sample, or in order to make the ratio comparable from one sample

to another. In this study 100 was used as the standard number of tokens, each language sample having been divided up into 30 consecutive 100-word segments. The TTR for each of these 100-word segments was then computed.

To determine, first, the internal consistency (i.e., how well a random half of the sample measures what the whole sample measures) of the 3,000-word sample for each subject, the *t*-test for related measures (10) was used. This was computed by dividing at random the 30 TTR's[3] for each subject into two sets and finding the group mean for each half. From this procedure there resulted two sets of ten means each for each group of subjects. Each set of ten means was averaged, giving two mean values for each group of subjects. The difference between these two mean values was evaluated. The value of *t* for the difference between the two means thus obtained for the schizophrenic patients was .219, and that for the freshmen was .430. As neither of these values of *t*, with nine degrees of freedom, is significant at the 5 per cent level of confidence it would appear that there is no reliable difference between the two means for each group, or that the internal consistency of the language samples is high.

A test of the hypothesis that there is no difference between the variances of the distributions of the SD's of individual samples of the schizophrenic patients and of the freshmen is afforded by the *F* test (10). It will be recalled that each individual sample is made up of 30 segments, for each of which a TTR was computed. When *F* was computed as the ratio of the variance of the distribution of the SD's for the schizophrenic patients to that for the freshmen, the value obtained was 2.2. Since the value of *F*, with nine and nine degrees of freedom, needed for significance at the 5 per cent point is 3.18, the hypothesis of no significant difference is tenable. That is to say, the TTR's of the schizophrenic patients did not vary more from segment to segment than did those of the freshmen.

Table 1 gives the distribution of the mean segmental TTR's for the individual freshmen and schizophrenics; each individual mean represents the average of the 30 segmental TTR's computed for each sample. This table indicates a tendency for the mean TTR to be generally lower in the case of the schizophrenics, only one freshman having a lower ratio than the patients with the highest ratios. It is to be noted, also, that the range for the schizophrenic group is much greater than for the freshman group, extending from .49 to .62 for the former, and from .61 to .67 for the latter.

The group mean TTR for the schizophrenics was .57, with a standard error of .0124, and that for the freshmen was .64, with a standard error of .0043. In order to test the significance of the difference between these two means the *t*-test (10) was applied. The value obtained for *t* was 5.61, which, with 18 degrees of freedom, is significant at the 1 per cent level of confidence. Therefore, the hypothesis that these two samples were drawn from populations whose means are equal may be rejected.

However, one of the assumptions underlying the *t*-test when used to test the significance of the difference between means of independent small samples is that the true variance of the population from which one sample is drawn must be equal (or approximately equal) to the

[3] As the language sample of one schizophrenic patient consisted of only 2,800 words, because he was withdrawn from the hospital before 3,000 words could be obtained, only 28 TTR's were obtained in his case.

TABLE I
Mean TTR's for the individual subjects ranked in descending order

Schizophrenic patients			Freshman subjects		
Mean TTR	S.D.	C.V.	Mean TTR	S.D.	C.V.
			.67	.056	8.36
			.66	.037	5.61
			.66	.035	5.30
			.64	.040	6.25
			.64	.057	8.91
			.64	.057	8.91
			.64	.053	8.28
			.63	.053	8.41
			.63	.042	6.67
.62	.048	7.74			
			.61	.057	9.34
.61	.044	7.21			
.60	.048	8.00			
.58	.050	8.62			
.57	.071	12.46			
.56	.030	8.93			
.56	.056	10.00			
.55	.064	11.64			
.53	.071	13.40			
.49	.066	13.47			

true variance of the population from which the other sample is drawn. In order to discover whether or not this assumption is valid in these samples the *F* test was applied. When *F* was computed as the ratio of the variance of the distribution of the mean TTR's for the schizophrenics to that of the freshmen, the value obtained was 8.36, which, with nine and nine degrees of freedom, is significant at the 1 per cent point. It might be possible to interpret this as invalidating the above use of the *t*-test with these data. There is doubt on this point, and while some statisticians might accept the *t*-test as here applied, it was thought best to treat the data in another and somewhat different way. Consequently, as a further check on the reliability of the difference between the two group means, *t* was used to set limiting values for each group outside of which any exact hypothesis as to the value of the true mean may be rejected with a given degree of confidence (10). At the 1 per cent level of confidence the limiting values of the true mean for the patients were .6085-.5277, and for the freshmen they were .6556-.6276. Since there is no overlap in these confidence intervals, we may be practically certain that the difference between the group mean TTR for the schizophrenics and that for the freshmen indicates a real difference between the two groups.

In general it may be concluded that the schizophrenic patients tended to have lower mean segmental TTR's than did the freshmen. In other words, the schizophrenic patients employed smaller vocabularies than did the freshmen.

Interpretation of these differences in regard to the TTR's of the schizophrenic patients and the freshmen must necessarily be made with caution because of several variables in the two groups, especially within the schizophrenic group, such as age, time of onset of illness, intellectual level and educational advantages, which the experimenter was not able to control rigidly within the limitations of this study. However, two possible relationships may be pointed out, namely, that existing between the intel-

lectual level and the TTR and that existing between certain clinical pictures presented by the patients and the TTR.

From a preliminary study by Zipf (17) in which he used a measure similar in some respects to the TTR, it may be inferred, although it cannot be stated conclusively, that the TTR probably correlates positively with mental age. When the schizophrenic patients are ranked according to their mean TTR and what estimates could be obtained of their original intellectual level, it would appear that a positive correlation would result.

Schizophrenics *Case*	*Type*	*Mean TTR*	*Estimates of Intelligence or Education*
7	paranoid	.62	"Very superior"
3	catatonic	.61	"High ave. to sup."
9	paranoid	.60	Eighth grade edu.
1	paranoid	.58	"Superior"
10	paranoid	.57	"Superior"
8	unclassified	.56	College grad.
6	hebephrenic	.56	"At least high ave."
5	paranoid	.55	"Ave. or above"
4	paranoid	.53	"Slightly below ave."
2	hebephrenic	.49	"Too deteriorated to estimate." Eighth grade education

Certainly the highest mean TTR was made by the schizophrenic with the highest intelligence, while the three lowest TTR's were made by the three patients with probably the lowest intelligence. Five other patients with probably high average to superior intelligence ranked in between.

No statements characterizing the various types of schizophrenia in terms of the TTR would be justified by the above tabulation.

Despite the probability that a positive correlation exists between the TTR and the intellectual level, the fact still remains that there were differences between the mean TTR's for the schizophrenics who ranked highest intellectually and most of the freshmen. As the TTR represents the relationship between the number of different words and the total number of words, the lower TTR's of the schizophrenics would obviously indicate a smaller number of different words used, hence more repetitions of the same words. Clinically, schizophrenic patients present a tendency to repetition of behavior known as stereotypy which may be of attitude, movement, or speech. When the same word, phrase or sentence is repeated the stereotypy is known as verbigeration (13). It is possible, then, that the lower mean TTR's for the patients represent to some degree in a quantitative manner this clinical picture of stereotypy.

2. *Gramatical analysis.* For this part of the study eight conventional parts of speech were used, namely, nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions and interjections. The articles were tabulated separately and then considered both alone and in conjunction with the adjectives. For the classification of words on this basis, the following rules were followed:

1. A noun used as an adjective was tabulated as an adjective only if the dictionary (14) gave the adjectival use as possible. For example, *family* in the combination *family prayers* was considered as an adjective as the dictionary gives this usage. However, the word *football* in the combination *football championship* was tabulated as a noun as nc

TABLE 2

Relative frequency of usage of the different parts of speech expressed as percentage of the total number of words used by the two groups, 29,800 in the case of the schizophrenic patients, and 30,000 in the case of the freshman subjects. The range values are from the individual samples

	Schizophrenic patients		Freshman subjects	
	%	Range	%	Range
Nouns	13.04	10.40–16.63	15.39	12.67–18.53
Pronouns	22.68	19.33–24.73	17.96	14.40–20.40
Verbs	26.28	24.27–30.47	22.95	20.50–24.47
Adverbs	11.54	7.00–17.97	10.16	8.87–11.20
Conjunctions	6.53	4.10– 8.77	8.83	7.33–11.40
Prepositions	7.48	4.30–10.00	10.00	8.80–11.00
Interjections	2.64	.53– 4.43	1.26	.47– 2.00
Adjectives	5.37	3.77– 7.10	6.69	5.67– 7.87
Articles	4.48	2.53– 6.87	6.79	5.27– 9.07
Adjs. and Arts.	9.85	8.60–12.40	13.48	11.43–16.40

adjectival use is mentioned in the dictionary.

2. Participles were classes as adjectives and gerunds as nouns only when this was indicated as permissible by the dictionary. Otherwise, they were classed with the verbs.

3. All pronouns were classified under pronouns whether modifying nouns or not.

4. The neologisms or coined words of the patients were interpreted according to the parts of speech that they seemed functionally to assume in the sentence; if in isolation, they were considered as nouns.

TABLE 3

Values of t and F obtained from testing significance of the difference in usage of certain grammatical categories, based on percentages of total sample, between schizophrenic patients and freshmen

	Values of t	Values of F
Adjectives	3.22	2.61
Adverbs	1.44	10.10
Nouns	2.50	1.53
Pronouns	5.30	1.42
Verbs	3.92	2.19
Adjs. and Articles	5.34	1.58
Articles	4.20	1.12
Prepositions	5.04	4.50
Conjunctions	3.43	1.44
Interjections	2.98	6.72

With 18 degrees of freedom, the values of t required for significance are: at the 1% level of confidence $t=2.88$; at the 5% level of confidence $t=2.10$.

With 9 and 9 degrees of freedom, the values of F required for significance are: at the 1% point $F=5.35$; at the 5% point $F=3.18$.

Table 2 gives the results of this grammatical analysis for the schizophrenics and freshmen, respectively. The t-test was applied to test the significance of the differences between the various percentages for the two groups, and the values of t obtained are given in Table 3. As can be seen, all of the t values thus obtained are significant at the 1 per cent level, except that for nouns which is significant at the 5 per cent level, and that for adverbs, which is not significant at that level. From this we may conclude that the schizophrenic patients used significantly fewer nouns, conjunctions, prepositions, adjectives, and articles than did the freshmen, and significantly more pronouns, verbs, and interjections.

The F ratio, involving the variances of the distributions of percentages (based on total words per sample) for each grammatical category for the two groups, resulted in the values of F also given in Table 3. Here the only significant results were with respect to the adverbs and interjections, which were significant at the 1 per cent point, and the prepositions

which was significant at the 5 per cent point. This would indicate that only in the use of adverbs, prepositions, and interjections did the schizophrenic patients show significantly greater variability than did the freshmen.

The ranges shown in Table 2 represent the highest and lowest percentage for each part of speech in the individual language sample for each subject, the total number of words being 3,000 in each instance, except for the one patient who had only 2,800. It will be noted that the schizophrenic patients showed a greater range for all parts of speech except the pronouns and the adjectives and articles combined, where the freshmen had a slightly greater range.

TABLE 4

Relative frequency of use of the different parts of speech expressed as percentage of the total number of words used by the two groups, 29,800 for the schizophrenic patients and 30,000 for the freshman subjects, compared with data from French, Carter, and Koenig (6) on telephone conversation.

	Tele phone Conv.	Schizo-phrenics	Fresh-men
Nouns	15.91	13.04	15.39
Pronouns	18.22	22.68	17.96
Verbs	22.39	26.28	22.95
Adjs. and advs.	10.06	16.91	16.85
Preps. and conjs.	12.62	14.01	18.83
Articles	5.60	4.48	6.79
Interjections	8.08	2.64	1.26

Table 4 shows the group percentages for each part of speech for the schizophrenics and for the freshmen, as compared with percentages computed from data given by French, Carter, and Koenig (6), in a study of telephone conversations. The data taken from this study were reorganized, wherever given in such form as to make it possible, in order to make them more nearly comparable to those of the present study. However, there were some differences in the French, Carter, and Koenig material that could not be changed so as to make it accord with that of the present study. For example, they classified all forms of *yes* and *no* under interjections, while such words were classed as adverbs in the present study, and they also classified laughter as an interjection, while it was ignored in this study. In addition, they grouped letters and numbers together under a separate heading, not classifying them as representing a part of speech, while letters were usually called nouns and cardinal numbers, adjectives, in this study. Therefore, this group of items, representing 5.05 per cent of the total number of words in their study, was ignored in the comparisons. These differences in procedure explain to some extent why the percentages of adjectives and of adverbs in the French, Carter, and Koenig data are considerably smaller than those for either of the two groups considered by the present experimenter, and why the percentage of interjections is considerably larger. However, it is interesting to note that the percentages for nouns, pronouns, and verbs in the French, Carter, and Koenig study approximate very closely the corresponding percentages for the freshman group used in this study, and hence are lower for pronouns and verbs than are those of the schizophrenic group, while the percentage of nouns is higher. In regard to prepositions and conjunctions the French, Carter, and Koenig percentage is lower than that for both the schizophrenic and freshman groups, but it more closely approximates that for the schizophrenic group. The percentage of articles in telephone conversation lies almost exactly half way between the percentage of articles for the schizophrenics and that for the freshmen.

Table 5 presents data from Horn (7)

TABLE 5

Relative frequency of use of the different parts of speech expressed as percentage of the total number of words used by the two groups, 29,800 for the schizophrenic patients and 30,000 for the freshman subjects, compared with data from Horn (7) on children

	Children		Schizophrenics		Freshmen	
	Mean	Range	Mean	Range	Mean	Range
Nouns	50.65	42.2–59.1	13.04	10.40–16.63	15.39	12.67–18.53
Pronouns	2.25	.9– 3.6	22.68	19.33–24.73	17.96	14.40–20.40
Verbs	27.75	16.9–38.6	26.28	24.27–30.47	22.95	20.50–24.47
Adverbs	5.65	2.5– 8.8	11.54	7.00–17.97	10.16	8.87–11.20
Conjunctions	1.5	.3– 2.7	6.53	4.10– 8.77	8.83	7.33–11.40
Prepositions	1.1	.6– 1.6	7.48	4.30–10.00	10.00	8.80–11.10
Interjections	.6	0 – 1.2	2.64	.53– 4.43	1.26	.47– 2.00
Adjectives	13.45	10.1–16.	9.85	8.60–12.40	13.48	11.43–16.40

showing the range of percentages on parts of speech that 11 investigators have found in children's language, as compared to the percentages for the schizophrenics and freshmen found in this study. For case of comparison the experimenter averaged these ranges, each study having been done on only one child. Here we immediately note some striking differences. The children used approximately three to four times as many nouns as either the schizophrenics or freshmen. They used eight to ten times fewer pronouns, about half as many adverbs, four to six times fewer conjunctions, seven to ten times fewer prepositions, two to four times fewer exclamations, about the same number of adjectives as did the freshmen (hence more than the schizophrenics), and about the same number of verbs as did the schizophrenics (hence fewer than did the freshmen). Again no conclusive comparisons can be made because of the probably varying procedures used in making the grammatical analyses.

If reference is made also to the French, Carter, and Koenig data, one might conclude that while the relative proportions of the various parts of speech change greatly from childhood to the adult level, the differences among various samples of adults are much smaller. Certainly there is no apparent tendency for the schizophrenic patients to regress toward the childhood level with respect to the general grammatical construction of their language, unless it might be in regard to more frequent use of verbs.

3. *Word frequencies.* Table 6 gives a list of the 100 most frequently used words for the schizophrenic patients and the freshmen, respectively, the list for the latter having those words which are common to both lists arranged in order of frequency, while the list for the schizophrenic patients has the words corresponding to those of the freshmen arranged in order of sequence regardless of frequency. The 21 words in each of the two groups not common to both lists are arranged at the bottom of the table in order of frequency. Several interesting differences between the list for the schizophrenic patients and that for the freshmen can be noted in regard to the frequencies for various words. For example, the schizophrenics used *not* almost twice as many times as did the freshmen. In addition, *no* and *never* occur in the list for the schizophrenics, while no other clearly negative words occur in the first 100 words for the freshmen. Hence, we have the schizophrenics using these negative words 1,087 times to 484 times for the freshmen, the former

TABLE 6

List of 100 words most frequently used by schizophrenics and freshmen. The first 79 words common to both lists are arranged in descending rank order according to frequency of usage by freshmen. The remaining 21 words not common to both lists are arranged in order of frequency for the two groups at the end of the table

	Freshmen			Schizophrenics		
	Word	Part of Speech	Freq.	Word	Part of Speech	Freq.
1.	the	art.	1140	the	art.	735
2.	and	conj.	1113	and	conj.	785
3.	I	pron.	924	I	pron.	2501
4.	a	art.	788	a	art.	356
5.	to	prep.	779	to	prep.	635
6.	is	verb	629	is	verb	580
7.	it	pron.	623	it	pron.	729
8.	of	prep.	612	of	prep.	416
9.	that	pron.	599	that	pron.	633
10.	you	pron.	562	you	pron.	392
11.	not	adv.	484	not	adv.	942
12.	in	prep.	396	in	prep.	266
13.	he (He)	pron.	347	he (He)	pron.	244
14.	that	conj.	327	that	conj.	172
15.	have	verb.	305	have	verb	339
16.	do	verb.	304	do	verb	638
17.	they	pron.	276	they	pron.	321
18.	well	interj.	271	well	interj.	565
19.	was	verb	270	was	verb	412
20.	are	verb	238	are	verb	136
21.	if	conj.	234	if	conj.	164
22.	she	pron.	220	she	pron.	127
23.	we	pron.	218	we	pron.	79
24.	but	conj.	211	but	conj.	173
25.	or	conj.	204	or	conj.	150
26.	just	adv.	177	just	adv.	190
27.	for	prep.	175	for	prep.	128
28.	there	adv.	168	there	adv.	163
29.	with	prep.	165	with	prep.	98
30.	would	verb.	164	would	verb	226
31.	had	verb	159	had	verb	212
32.	what (uh?)	pron.	155	what (uh?)	pron.	297
33.	very	adv.	154	very	adv.	46
34.	think	verb	147	think	verb	131
35.	oh	interj.	143	oh	interj.	125
36.	about	prep.	141	about	prep.	133
37.	know	verb	139	know	verb	496
38.	on	prep.	138	on	prep.	109
39.	get	verb	125	get	verb	120
40.	at	prep.	117	at	prep.	73
41.	out	adv.	115	out	adv.	95
42.	will	verb	113	will	verb	52
43.	people	noun	111	people	noun	74
44.	something	noun	108	something	noun	86
45.	them	pron.	108	them	pron.	66
46.	this	pron.	100	this	pron.	85
47.	one	pron.	99	one	pron.	72
48.	me	pron.	96	me	pron.	272
49.	up	adv.	93	up	adv.	71
50.	when	conj.	93	when	conj.	114
51.	might	verb	89	might	verb	47
52.	then	adv.	85	then	adv.	100
53.	as	conj.	84	as	conj.	46
54.	things	noun	84	things	noun	80
55.	time	noun	83	time	noun	61
56.	because	conj.	82	because	conj.	149
57.	can	verb	78	can	verb	75

TABLE 6 (Continued)

	Freshmen			Schizophrenics		
	Word	Part of Speech	Freq.	Word	Part of Speech	Freq.
58.	were	verb	76	were	verb	58
59.	say	verb	75	say	verb	104
60.	good	adj.	75	good	adj.	47
61.	him (Him)	pron.	74	him (Him)	pron.	57
62.	go	verb	71	go	verb	56
63.	my	pron.	71	my	pron.	286
64.	cannot	verb	70	cannot	verb	93
65.	did	verb	70	did	verb	158
66.	like	prep.	69	like	prep.	82
67.	all	adj.	68	all	adj.	53
68.	so	adv.	62	so	adv.	75
69.	see	verb	62	see	verb	47
70.	am	verb	61	am	verb	167
71.	one	adj.	59	one	adj.	64
72.	some	adj.	59	some	adj.	54
73.	anything	pron.	59	anything	pron.	110
74.	could	verb	58	could	verb	121
75.	got	verb	56	got	verb	72
76.	want	verb	52	want	verb	62
77.	been	verb	52	been	verb	67
78.	way	noun	48	way	noun	58
79.	means	verb	48	means	verb	93
80.	his (His)	pron.	121	yes (uh huh)	adv.	173
81.	person	noun	118	be	verb	145
82.	an	art.	103	said	verb	109
83.	has	verb	102	no (hunh uh)	adv.	96
84.	who	pron.	102	why	interj.	89
85.	her	pron.	76	suppose	verb	85
86.	so	conj.	74	now	adv.	82
87.	by	prep.	71	guess	verb	73
88.	let	noun	65	here	adv.	73
89.	from	prep.	65	any	adj.	71
90.	other	adj.	63	thought	verb	70
91.	example	noun	63	mean	verb	66
92.	going	verb	62	sir	noun	65
93.	quite	adv.	61	thing	noun	56
94.	your	pron.	58	too	adv.	56
95.	which	pron.	57	all	noun	53
96.	does	verb	54	never	adv.	49
97.	always	adv.	54	understand	verb	49
98.	us	pron.	50	little	adj.	47
99.	then	conj.	49	right	adj.	45
100.	course	noun	48	tell	verb	43

group using them about two and one-half times more than the freshmen, when only the 100 most frequently used words are considered. Instead of the *never* used by the schizophrenics ,the freshmen used *always* about an equal number of times. Another interesting item is that the freshmen used *very* over three times as often as did the schizophrenics. When the verbs among these 100 most frequently used words for the group were considered, it was found that the schizophrenic patients used eight past tense verbs a total of 1158 times while the freshmen used six such verbs only 683 times. It is interesting to note, also, that two verbs carrying the connotation of indecision, *suppose* and *guess,* occur among the 100 most frequently used words of the schizophrenics for a total of 158 times, while no such words occur in the comparable list for the freshmen.

A more detailed comparison of frequencies for various words used by the schizophrenic group and by the freshman group possibly would show several interesting and differential facts. A consideration of the qualitative aspects of some of the words used by the two groups would also provide interesting material.

Because of the tendency shown in the TTR analysis for the schizophrenic patients to repeat words more frequently than do the freshmen, Fig. 1 is presented to show what proportion the 100 most frequently used words constituted of the total number of words for the two groups. We may refer to this as proportional vocabulary. The frequencies for each consecutive five words, starting with the most frequently used word, were added cumulatively for each group, and these successive cumulative frequencies were expressed as fractions of the total number of words. The curves show that the patients consistently use a smaller number of different words to represent any given percentage of the total number of words. For example, the schizophrenic group use only 33 words to make up 50 per cent of the total number of words, while the freshman group use 46 words to arrive at the same percentage. The entire 100 most frequent words constitute 68.32 per cent of the total number of words for the schizophrenics as a group, and 62.91 per cent for the freshmen. Superimposed on these curves is a similar curve taken from the French, Carter, and Koenig (6) study, indicating that the 100 most frequently used words in the telephone conversations analyzed formed 75 per cent of the total number of words. The curve on written material was also given by French, Carter, and Koenig, and was taken by them from Dewey (5). According to it, the 100 most frequently used words in written material form only 56 per cent of the total number of words used. A consideration of all four curves shows that the telephone conversation and written English represent the extremes in this factor of repetition, or number of types constitut-

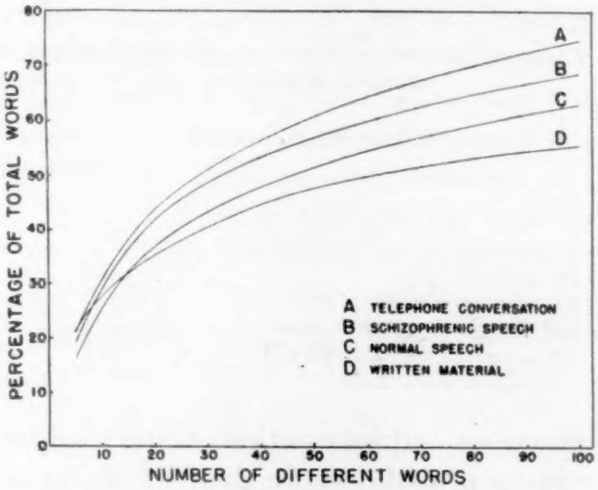


FIG. 1. Curves showing the cumulative percentages of the total words for the 100 most frequently used words. A, telephone conversation (0); B, schizophrenic subjects; C, freshman subjects; D, written material (5).

ing a given percentage of the total number of tokens, the telephone conversation being the most repetitious and the written English the least. This might be expected from the stereotyped, truncated nature of telephone conversation as compared to the reflective style of written English in which a premium is placed on variety. That the curves for the two groups considered in this study should fall in between these extremes, that for the schizophrenic group more nearly approximating that for the telephone conversation and the curve for the freshman group being closer to the one for the written English, might also be expected, considering the repetitious nature of the schizophrenic speech. The freshmen appear to have been more successful in introducing variety and flexibility into their spoken language. The lower end

of the curves indicates the interesting fact that the curve for written English overlaps the other three and is higher for the first five words or so. This might be explained as being due to the completeness of written English as compared to conversations, the articles and the conjunctions and prepositions being used by writers probably more than by speakers. An examination of Dewey's list shows that the five most frequently used words in the written English samples were *the, of, and, to,* and *a.*

Another analysis that suggests itself, because of the schizophrenic's self-preoccupation and his tendency to ignore his environment, is the relative frequency of referrals to self and of referrals to others found in the language of the two groups. This analysis was made by computing the percentage of the total number of words represented by the different personal pronouns. Table 7 shows the results of this computation. The most striking fact in this table is that references to self, using some form of the first person singular pronoun (*I, my, mine, me, myself*), make up 10.42 per cent of the total number of words for the schizophrenic group, while they represent only 3.69 of the total for the freshmen. On the other hand the freshmen use the first person plural pronoun in its various forms (*we, our, ours, us, ourselves*) three times as often as do the schizophrenics; the second person pronoun, plural and singular (*you, your, yours, yourself, thee, thou*), almost twice as frequently as do the schizophrenics; and the third person pronoun, singular and plural (*he, his, him, himself, she, her hers, herself, it, its, itself, they, their, theirs, them*), almost 20 per cent more often.

TABLE 7

Relative frequency of use of the different personal pronouns expressed as percentage of the total number of words for the two groups, 29,800 for the schizophrenic patients and 30,000 for the freshmen

	Schizophrenics		Freshmen	
	N	%	N	%
1st person sing.	3104	10.42	1107	3.69
1st person plural	102	.32	315	1:05
2nd person sing. and plural	429	1.44	643	2.14
3rd person sing. and plural	1645	5.52	1923	6.41

The schizophrenic patients used a total of 14 neologisms, or coined words. These words are *banoon, d-s, d-t, dokey, g-m, g-o-d-t, okey-dokey, oke, pard, p-a-r-d, recognization, strob, striked, woozy, adjects.* Neologisms were not found in the freshmen samples.

IV. SUMMARY AND CONCLUSIONS

Three-thousand-word language samples were obtained from each of ten schizophrenic patients, five males and five females, and ten University of Iowa freshmen, four males and six females, the latter ranking above the 90th percentile on the Composite Score of the Iowa Qualifying and Placement Examinations. An interview situation was employed, involving the interpretation of 14 proverbs, the interviews being recorded by an electric dictaphone technique without the subjects' knowledge.

A word count was then made for each language sample, each word being tabu-

lated according to its frequency in consecutive 100-word segments and according to its grammatical usage. Three types of analysis were made: (1) the type-token ratio, computed by dividing the number of different words in each 100-word segment by the total 100 words; (2) grammatical analysis; and (3) word frequencies.

1. When the *t*-test for related measures was applied to the language samples for both groups by dividing at random the 30 TTR's for each subject into two sets and finding the group mean for each half, it was found there was no significant difference between the two means either for the schizophrenics or for the freshmen.

2. When the ratio of the variance of the distribution of the standard deviations of the segmental TTR's for the schizophrenic patients to that of the freshmen was computed, the resulting *F* value was not significant, indicating that the schizophrenic patients did not vary more from segment to segment than did the freshmen.

3. The mean TTR's of the schizophrenic patients were generally lower than were those of the freshmen, and the range for the patients was much greater.

4. The group mean TTR of the schizophrenic patients was found to be significantly lower than the group mean TTR for the freshmen.

5. It is probable that a positive correlation exists between the TTR and the intellectual level, according to previously reported findings, and judging by the indicated relationship between the TTR's of the patients and their probable intellectual levels when both were ranked in descending order. However, there were differences between the mean TTR's for the schizophrenic patients who ranked highest intellectually and most of the freshmen.

6. When the *t*-test was applied to test the difference between the two groups in terms of the relative frequency of usage of the eight grammatical categories, expressed as percentages of the total number of words used, it was found that the schizophrenic patients used significantly fewer nouns, conjunctions, prepositions, adjectives, and articles than did the freshmen, and significantly more pronouns, verbs, and interjections.

7. The *F* ratio, involving the variances of the distributions of percentages (based on total words per sample) for each grammatical category for the two groups revealed that the schizophrenic patients showed significantly greater variability than did the freshmen in the use of adverbs, prepositions, and interjections.

8. Comparison of the relative proportions of the various parts of speech found in this study with those given in another study on telephone conversation, for presumably normal adults, indicates a very close approximation between the percentages of nouns, pronouns, and verbs used in telephone conversation and those used by the freshman group. The procedure used in the former study for classifying these three parts of speech was quite similar to that used in the present study. The procedures for classifying the prepositions and conjunctions, and the articles also apparently were similar, but the percentage for the former was considerably lower for the telephone conversation than for the freshman language, and the percentage of articles was slightly lower. The percentages of adjectives and adverbs were also considerably lower for the telephone conversation than for either the schizophrenic or freshman samples, and the

percentage of interjections was a great deal higher, but the procedures for the classification of these two groups of words differed considerably in the two studies.

The most definite differences between the schizophrenic patients and the normal adults in this and the other study lie in the fact that the patients used proportionately more pronouns and verbs, and proportionately fewer nouns and articles.

9. A general comparison with similar data on children under six and one-half years of age showed several marked differences between the percentages on the parts of speech for the children and those for the two groups in this study. The children used many more nouns and many fewer pronouns, adverbs, conjunctions, prepositions and interjections than either the schizophrenic or freshman group. In the percentage of verbs the children more closely resembled the schizophrenic group, and their percentages of adjectives was nearly the same as for the freshman group.

10. Assuming that the probably different procedures in the grammatical analyses of the three studies permit general comparisons, it would appear that while the relative proportions of the various parts of speech change greatly from childhood to the adult level, the differences among various samples of adults are much smaller. There was little evidence from this analysis that schizophrenia constitutes a regressive tendency, except for the more frequent use of verbs, the other findings for the children and schizophrenics, respectively, being decidedly different.

11. When a list of the 100 words most frequently used by the schizophrenics and by the freshmen was made, it was found each list had 21 words not common to the other.

12. The total frequencies for these 100 most frequently used words constituted 68.32 per cent of the total number of words used by the schizophrenics and 62.91 per cent of those used by the freshmen, the schizophrenics consistently using a smaller number of different words to make up any given percentage of the total up to this figure. For the schizophrenics 33 different words (types) constituted 50 per cent of their total sample of 29,800 words (tokens); for the normals 46 types constituted 50 per cent of their 30,000 tokens.

13. A comparison of the relative proportion of referrals to self and referrals to others, as indicated by the use of personal pronouns by the two groups, shows that the schizophrenics used more first person singular pronouns, and fewer first person plural, second person plural and singular, and third person plural and singular pronouns than did the freshmen. *I, my, mine, me,* and *myself* represented 10.42 per cent of the total number of words for the schizophrenics, and only 3.69 per cent of the total words for the freshmen.

14. Several interesting differences in the frequencies of occurrence of specific words among these 100 most frequently used words were noted, such as the fact that negative words (*not, no,* and *never*) have a frequency two and one-half times larger in the schizophrenic list than in the freshman list, and that verbs in the past tense had a frequency a little less than twice as large in the schizophrenic list as in the freshman list.

The conclusion can be stated that the measures used do make possible the quantitative expression of certain differences among samples of spoken lan-

guage. Statistically significant differences between schizophrenic language and the language of superior university freshmen, as these types of language were here sampled, were indicated by the measures of vocabulary extent and 'flexibility', and of grammatical structure. The measures of word frequency were also suggestive of some possibly important differences between the two groups.

These findings are to be evaluated with clear awareness that they may not be due entirely to the schizophrenia, since, as was explained in the Introduction, there were necessarily differences between the two groups with regard, particularly, to intelligence and scholastic training, and the relevance of these differences cannot, at this stage of investigation, be clearly judged. The degree to which such 'intellectual' factors are related to the language measures employed is not yet known; and the problem of measuring the intelligence of psychopathological individuals is by no means simple. Insofar as any conclusions may be drawn about 'schizophrenic language' on the basis of this study, they would appear to suggest the possibility that such language differs from the language of 'normal' persons in being (a) less highly differentiated in structure—the ratio of different words (types) to total words (tokens) is lower, as shown by the analyses in terms of type-token ratio and proportional vocabulary; (b) more negatively toned; (c) indicative of preoccupation with the past, as shown by relatively more past tense verbs; (d) indicative of more self-reference, as shown by more frequent occurrence of self-reference terms in the first person singular pronoun class; (e) characterized by a slight tendency toward the use of neologisms; (f) featured by a probable peculiarity of grammatical structure, represented by relatively more pronouns and verbs and fewer nouns and articles, which might possibly be suggestive of excessive self-preoccupation and 'instability'. Moreover, such comparison as could be made of the 'schizophrenic language' and that of children (7) provided little ground for the view that schizophrenia constitutes a regression to childhood behavior patterns, in that the language of the schizophrenics, as measured, bore no striking resemblance to that of the children, except possibly in the proportionate number of verbs.

Again, it is to be emphasized that this study was designed primarily to explore the possibilities of language measurement. From this point of view, its results may be regarded as definitely promising. Any conclusions concerning the nature of 'schizophrenic language' are advanced only for their suggestive value.

BIBLIOGRAPHY

1. Betts, E. A. An experimental appraisal of certain techniques for the study of oral composition, Ph. D. Diss., Univ. Iowa, 1931.
2. Cameron, N. A study of thinking in senile deterioration and schizophrenic disorganization, *Amer. J. Psychol.*, 1938, *51*, 650-655.
3. Cameron, N. Reasoning, regression and communication in schizophrenics, *Psychol. Monogr.*, 1938, *50*, No. 1, 1-34.
4. Carroll, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution, *Psychol. Rec.*, 1938, 2, 379-386.
5. Dewey, G. Relative frequency of English speech sounds, *Harvard Studies in Educ.*, IV, Harvard University Press, 1923.
6. French, N. R., Carter, C. W., Jr., & Koenig, W., Jr. The words and sounds of telephone conversation. *Bell System Tech. J.*, 1930, *9*, 290-324.

7. Horn, M. D. An investigation of the vocabulary of kindergarten children, M.A. Thesis, Univ. Iowa, 1927.
8. Johnson, W., *Language and Speech Hygiene,* Chicago: Institute of General Semantics, 1939, 10-12.
9. Johnson, W., Studies in language behavior: I. A program of research, *Psychol. Monogr.* 1944, *56,* 1-15.
10. Lindquist, E. F. *Statistical Analyses in Educational Research.* New York: Houghton Mifflin Co., 1940.
11. Mann, M. Studies in language behavior: III. The quantitative differentiation of samples of written language, *Psychol. Monogr.,* 1944, *56,* 41-74.
12. Mitchell, M. Unpublished research.
13. Noyes, A. P. *Modern Clinical Psychiatry,* Philadelphia: W. B. Saunders Co., 1934, 103.
14. *Webster's New International Dictionary* (2nd ed.), Springfield, Mass.: G. and C. Merriam Co., 1934.
15. White, W. A. The language of schizophrenia, *Arch. Neurol. Psychiat.,* 1926, *16,* 395-413.
16. Woods, W. Language study in schizophrenia, *J. Nerv. Ment. Dis.,* 1938, *87,* 290-316.
17. Zipf, G. K. Observations of the possible effect of mental age upon the frequency-distribution of words, from the viewpoint of dynamic philology, *J. Psychol.,* 1937, *4,* 239-244.

III. THE QUANTITATIVE DIFFERENTIATION OF SAMPLES OF WRITTEN LANGUAGE

III. The Quantitative Differentiation of Samples of Written Language[1]

MARY BACHMAN MANN
University of Iowa

I. INTRODUCTION

This study is part of a previously described program of research concerned with the general problem of language behavior (16).

The present investigation is concerned primarily with the objective of developing reliable and differentiating measures of language behavior, and, to a limited extent, with determining the intercorrelation of the measures, their relation to other pertinent variables, and with indicating the normal characteristics of language behavior as contrasted with disorder in such behavior.

The scientific study of language behavior has been carried on by many investigators, among them Piaget (18), Cameron (5), Thorndike (22), Horn (13), Zipf (25), Carroll (10), Skinner (20), Jersild and Ritzman (14), Balken and Masserman (2), and Fairbanks (12), to mention only a few.[2]

None of the previous investigators has been precisely concerned with the particular issues around which the present study is centered. In the first place, the present study is strictly quantitative, and this fact serves to differentiate it from a considerable proportion of previous investigations of language. Secondly, this study is concerned with the language behavior of specified individuals, a fact which differentiates it from practically all of the word-frequency studies such as those of Thorndike and Horn, in which large samples of language drawn from a variety of sources were studied but with no attention given to the characteristics of the language of individuals. Thirdly, some of the measures used in the present study, particularly the type-token ratios, have not been employed, as they are here used, in any previous studies with the exception of the one by Fairbanks (12) which may be regarded as a companion study to this one.

What was desired, for purposes of this particular study, was a sampling of the language of persons who could be regarded definitely as psychopathological, but who could nevertheless produce written language, and a sampling of the language of persons who could be regarded as definitely superior in verbal ability, but who might not be regarded as 'verbal specialists', such as outstanding novelists, scientific writers, etc. The study is concerned, first of all, with the specific problem of whether and in what respects 'adequate' and 'inadequate' language might be differentiated *quantitatively*. The problem of ascertaining the particular factors responsible for any demonstrated differences between the adequate and inadequate language is secondary to the main investigation, but it has been considered in some degree.

---

[1] This study was done in the Department of Psychology at the State University of Iowa as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The study was directed by Wendell Johnson, and is part of a program of research on language behavior. The writer is grateful to Dr. Andrew H. Woods, Director, and the staff of the Iowa State Psychopathic Hospital, and Dr. Leonard P. Ristine, Superintendent, and the staff of the Mt. Pleasant State Hospital, for their cooperation in securing subjects for the investigation.

[2] The reader who is interested in the study of language from the standpoint of vocabulary and word lists will find an excellent summary in Fries, Charles C. and Traver, A. Aileen, *English Word Lists,* American Council on Education, Washington, D.C., 1940, pp. 109.

Adults schizophrenic patients were selected as the subjects from whom samples of 'inadequate' language were to be obtained. Other specific types of subjects might have been chosen; subjects might have been selected, for example, solely on the basis of educational level, or of intelligence test scores. Aphasics might have been used in order to obtain 'inadequate' language. Aphasics, however, might be expected to produce language 'inadequate' in some relatively rare sense. And schizophrenics were preferred to persons mainly characterized by low-grade test-intelligence, or by low scholastic achievement, because insofar as their language is 'inadequate' it would appear to be so in a peculiarly significant sense from the standpoint of social adjustment. Thus, in the case of schizophrenics, neuro-linguistic inadequacy, insofar as it may exist, may reasonably be judged to have a significance beyond that of the neuro-linguistic inadequacy involved in 'simple' low-grade 'intelligence'.

Having selected the subjects from whom the 'inadequate' language samples were to be obtained, the problem of selecting a contrasting group of subjects presented itself. This problem was essentially that of selecting subjects from whom relatively high-grade language behavior might be expected, but who could be counted upon not to produce language that was highly 'adequate' in some relatively exceptional respect. Superior 'literary' language, for example, was to be avoided. After due consideration, the decision was made to select subjects who were not noted as being talented in some exceptional linguistic respect, who were behaviorally and socially normal in the sense, at least, that they could function as freshmen in a large university, and who were neuro-linguistically superior in the sense that they scored relatively very high on a battery of largely verbal tests administered to them on the occasion of their entering the university which they were attending.

The question might be raised as to the advisability of selecting 'normal' subjects matched with the psychotic patients with respect to such factors as 'intelligence', educational status, etc. The most important consideration in this connection is simply that such a procedure would probably have militated against the main purpose of the study, in that it would have made less likely the obtaining of two definitely differing samples of language. It was a primary consideration that two such samples be obtained if the problem of the quantitative differentiation of language samples was to be fruitfully investigated. A determination of the respects in which language samples of the type were utilized might be quantitatively differentiated would appear to be basic to any study of the relation of specific factors, such as 'intelligence', for example, to measurable aspects of language behavior.[3]

The language obtained from the psychotic subjects used in this investigation definitely constitutes a sample of the

[3] The question as to whether the language of schizophrenics differs, insofar as it does, from the language of superior university freshmen, because the schizophrenics are less "intelligent," raises an extremely complicated issue. It is not to be ligthtly dismissed, for example, that the phrase "highly intelligent schizophrenic" may be in a basic sense self-contradictory. The fact that an "intelligence test" shows a schizophrenic to be superior mentally probably tells, from one point of view, as much about the test as it does about the patient. The schizophrenic offers a means of validating the test quite as definitely as the test offers a means of evaluating the patient. A particularly pertinent answer to the test on which a schizophrenic scores a high "intelligence quotient" is that, when all is said, the schizophrenic is in custody. The issue is not a simple one by any means; further discussion of it, however, is hardly relevant to the present purposes.

language of schizophrenics, but whether its differentiating characteristics are due to 'schizophrenia' is a question, not without interest, but not of primary concern in this study. Of course, insofar as the differentiating characteristics of the schizophrenics' language cannot be attributed to something else, it would appear reasonable to regard them as due to, or as involved in, whatever may be designated by the term 'schizophrenia'. The relation of test-intelligence and educational level, at least, as well as that of sex, to the quantitatively expressible aspects of the schizophrenics' language has been ascertained to some extent in the present investigation. It is to be clearly understood that one is to be cautious, though not to the point of impotence, in drawing from this study any generalizations concerning 'the language of schizophrenia', since the study is designed primarily to yield generalizations with respect to another problem, namely that concerning the quantitative differentiation of samples of written language.

II. STATEMENT OF THE PROBLEM

This study is concerned with the following specific problem: the quantitative differentiation of samples of presumably adequate and inadequate written language, as obtained from superior university freshmen and schizophrenic patients, respectively, in terms of the following specific measures:

(1) The ratio of types (different words) to tokens (total words used).
(2) The relative frequency of usage of certain grammatical categories.
(3) The ratios of the frequency of occurrence of adjectives to verbs, adjectives to nouns, and adverbs to verbs, respectively.
(4) The relative frequency of specific types, expressed as percentage of tokens.

III. PROCEDURE

Two groups of adults served as subjects in this investigation: (1) twenty-four psychotic patients diagnosed as schizophrenic were selected to represent a group presenting psychopathological or inadequate language; (2) twenty-four superior university freshmen were selected to represent a group presenting relatively adequate language. A summary of the main characteristics of these two groups follows.

At the time the data were secured the patients were all confined in the Mt. Pleasant State Hospital at Mt. Pleasant, Iowa. Thirteen of the twenty-four had been previously examined at Iowa State Psychopathic Hospital, Iowa City, and the diagnosis of schizophrenia made by the psychiatrists at the Iowa State Psychopathic Hospital had been confirmed by the staff at the Mt. Pleasant State Hospital. These particular schizophrenic patients were selected because of the relatively maximum certainty of the diagnosis, and the possibility of securing their cooperation in the proposed writing situation. The patients, twelve male and twelve female, ranged in age from sixteen to forty-nine years, with an average age of thirty-two years; four (one male and three females) have been married. The average duration of present confinement in the Mt. Pleasant State Hospital prior to their service as subjects for this investigation was three years and three months, the range being from one year to eight years. The average duration of the illness, taken from the time of the first psychotic symptoms, as shown in the patient's hospital record,[4] was five and one-half years, ranging

[4] This must be considered as only an indication of duration of illness since it is difficult if not impossible in many cases to determine when the illness began. In the case of the disease

from one year to eleven years. Prior to their commitment in the hospital the patients had been engaged in the following occupations: laborer, accountant, farm laborer, high school student, college student, university law student, button cutter, pharmacy clerk, school teacher, telephone operator, hospital maid, and housewife. The level of educational attainment ranged from grade eight to college graduate; sixteen of the twenty-four were high school graduates and ten of those sixteen had some college training. Of the fifteen patients for whom intelligence ratings were available, the range in I.Q. points was from 78 to 138, the mean I.Q. being 99. It should be pointed out that mere I.Q. scores on these patients have little meaning, and care must be exercised in interpreting such scores. Where it was possible to do so, a vocabulary[5] score, or a verbal scale I.Q., and a performance scale I.Q. have been given. The intelligence tests were all administered by the hospital psychometrist and judgments as to probable classification are those of the psychometrist. Of the tests used, two were Wechsler-Bellevue Adult Scale; ten were Revised Stanford-Binet, Form L; one was Revised Stanford-Binet, Form M; and two were the 1916 Stanford Revision of the Binet-Simon Test.

Within the diagnosis of schizophrenia, twelve patients had been further classified as hebephrenic, three as simplex, seven as paranoid, and two as catatonic. The following abstracts present data concerning the individual patients. The information contained in these abstracts was taken from the hospital records for each patient.

Case 1. Diagnosis: Schizophrenia, hebephrenic type. A married female, thirty-eight years of age; completed college education and taught school one year after graduation before being married. First psychotic symptoms in 1931, present commitment to Mt. Pleasant State Hospital began in 1932, having previously been institutionalized in private sanitariums on two occasions. Scored Intelligence Quotient of 107 on Wechsler-Bellevue Adult Scale, Verbal Scale I.Q. 112, and Performance Scale I.Q. 100. Classification by psychometrist: Above Average; some inefficiency.

Case 2. Diagnosis: Schizophrenia, hebephrenic type. A single male, thirty years of age; educated through tenth grade in high school; occupation before committed to hospital, none. First psychotic symptoms one year before commitment to Mt. Pleasant State Hospital in 1936. No previous commitments. No intelligence test results available.

Case 3. Diagnosis: Schizophrenia, catatonic type. A single male, thirty years of age; educated through high school; previous occupations, working in restaurant and drug store. First psychotic symptoms in 1929, present commitment to Mt. Pleasant State Hospital began in 1936, having been committed previously for short periods in 1929 and again in 1931. No intelligence test results available.

Case 4. Diagnosis: Schizophrenia, paranoid type. A single female, aged twenty-seven years, educated through two years of college. First mental symptoms in 1930, present episode began in 1938, hos-

schizophrenia the onset is insidious, often extending over a period of several years before definite psychotic symptoms appear and are diagnosed. Furthermore, it is frequently difficult to ascertain from hospital records who first considered the behavior abnormal and diagnosed it as psychotic.

[5] According to Babcock (1) vocabulary is the best measure of the original intellectual level of the psychotic individual.

pitalized at Iowa State Psychopathic in 1939, then committed to Mt. Pleasant State Hospital and confined there since. Scored Intelligence Quotient of 138 on Revised Stanford-Binet, Form L, passing vocabulary at Superior Adult III level. Psychometrist commented that intellectual level was "very superior".

Case 5. Diagnosis: Schizophrenia, hebephrenic type. A single female, forty-eight years of age, graduated from high school and attended a teachers college one summer. Taught school three years before first attack which was in 1916 lasting for approximately one year, then did housework at home. Second and present attack began in 1938 when she was committed to Mt. Pleasant State Hospital. Scored Intelligence Quotient of 91 on Wechsler-Bellevue Adult Intelligence Scale, Verbal Scale I.Q. 106, Performance Scale I.Q. 78. Psychometrist's statement: "The patient's intellectual development is average."

Case 6. Diagnosis: Schizophrenia, paranoid type. A single male, thirty-two years of age, graduated from high school. Worked as a laborer prior to commitment to Mt. Pleasant State Hospital in 1938. The onset of the present episode was gradual, believed to have begun six or seven years before time of commitment. Scored Intelligence Quotient of 111 on Revised Stanford-Binet, Form L; vocabulary score, high average. Psychometrist's statement: "There is nothing remarkable about his performance. It was consistently good and warrants a classification of High Average Adult."

Case 7. Diagnosis: Schizophrenia, simple type. Single female, twenty-nine years of age, graduated from high school and worked successfully as a telephone operator for six years. First psychotic symptoms manifested in 1934, committed to Mt. Pleasant State Hospital in 1937 and confined there continuously since. Scored Intelligence Quotient of 78 on 1916 Revision of Stanford-Binet Test. Believed by psychometrist to reveal marked deterioration from an average intellectual development.

Case 8. Diagnosis: Schizophrenia, paranoid type. A single male, thirty-two years of age, educated through tenth grade at fifteen years, worked as a laborer prior to commitment in Mt. Pleasant State Hospital in 1939. First psychotic symptoms twelve to eighteen months before commitment and had spent six weeks in a private sanitarium. Scored Intelligence Quotient of 101 on the Revised Stanford-Binet, Form L, passed vocabulary test at Average Adult level. Classification by psychometrist: Average.

Case 9. Diagnosis: Schizophrenia, hebephrenic. A single female, eighteen years of age; present mental episode occurred during senior year of high school; confined to Mt. Pleasant State Hospital in 1939. Scored Intelligence Quotient of 108 on Revised Stanford-Binet, Form M. In this she showed a superior vocabulary. Classification: Average.

Case 10. Diagnosis: Schizophrenia, catatonic type. A single male, thirty-two years of age. Graduated from high school and attended college two and one-half years; confined to Mt. Pleasant State Hospital in 1932. Scored Intelligence Quotient of 101 on Revised Stanford-Binet, Form L. Classification: Average.

Case 11. Diagnosis: Schizophrenia, paranoid type. A married male, aged thirty-three years; educated through high school and two and one-half years of college; occupation, accountant. First symptoms in 1934, confined to Mt. Pleasant State Hospital since early in 1939. Scored Intelligence Quotient of 108 on

1916 Stanford Revision of the Binet-Simon Test. Classification: Average.

Case 12. Diagnosis: Schizophrenia, hebephrenic type. A single female, twenty-three years of age; educated through high school, and one semester of college. First psychotic symptoms in 1935, confined to Mt. Pleasant State Hospital continuously since June, 1938. Record of a Stanford-Binet Test given in 1927 with a C.A. of 11 indicated an I.Q. of 136; Stanford Revision of the Binet-Simon Test, Form L, administered in 1938 yielded an I.Q. of 97, vocabulary, 84. Psychometrist commented: "Vocabulary indicates a previous very superior level."

Case 13. Diagnosis: Schizophrenia, hebephrenic type. A single male, sixteen years of age. Psychotic symptoms began while he was in the ninth grade in high school. Committed to Mt. Pleasant State Hospital in 1939. Test results: Wechsler-Bellevue Adult Intelligence Scale, I.Q. 77, Verbal Scale I.Q. 65, Performance Scale I.Q. 95; Revised Stanford-Binet, Form L, I.Q. 84, vocabulary Average Adult. Classification: Formerly average—low average—poor school achievement.

Case 14. Diagnosis: Schizophrenia, hebephrenic type. A single female, aged thirty-six years; educated through high school and two years of college. First psychotic symptoms in 1932, hospitalized at Iowa State Psychopathic Hospital in 1933, then committed to Mt. Pleasant State Hospital and there since. Scored Intelligence Quotient of 83 on Revised Stanford-Binet, Form L, passing vocabulary test at Superior Adult II level. Judged by psychometrist to have originally been "at least high average".

Case 15. Diagnosis: Schizophrenia, simple type. A single male, twenty-five years of age, educated through eighth grade. Worked as a farm laborer. First psychotic symptoms appeared one month before commitment to Mt. Pleasant State Hospital in 1939. No intelligence test results available.

Case 16. Diagnosis: Schizophrenia, simple type. A single male, twenty-seven years of age; educated through high school and two years of junior college, then entered University Law School, where he was a good average student. First symptoms in 1934, at which time he was examined at Iowa State Psychopathic Hospital. Confined at Mt. Pleasant State Hospital since 1938. Results of the Revised Stanford-Binet Test, Form L, given in 1938, show an Intelligence Quotient of 87, vocabulary test high. Classification: Average. Shows deterioration from a probably superior intellectual development.

Case 17. Diagnosis: Schizophrenia, paranoid type. A single male, forty-nine years of age; educated through eighth grade. Occupation had been button cutter. First psychotic symptoms in 1930 when he was committed in Mt. Pleasant State Hospital for a short time, released, and then re-committed in 1936. No intelligence test results available.

Case 18. Diagnosis: Schizophrenia, hebephrenic type. A married female, thirty-eight years of age, educated through high school and two summer sessions at college. Taught school before and after marriage. First psychotic symptoms four months before commitment to Mt. Pleasant State Hospital in 1935. No intelligence test results available.

Case 19. Diagnosis: Schizophrenia, hebephrenic type. A single male, twenty-one years of age, educated through high school. Worked as a laborer before committed to hospital. Examined at Iowa State Psychopathic Hospital and hospitalized for seven months in 1936. Committed to Mt. Pleasant State Hospital in 1939. Scored Intelligence Quotient of

101 on Revised Stanford-Binet Test, Form L. Classification: Average.

Case 20. Diagnosis: Schizophrenia, hebephrenic type. A single female, forty-one years of age, educated through high school and two years at junior college. First psychotic symptoms in 1930, confined to Mt. Pleasant State Hospital for eight months. Re-entered the same hospital in 1934 and confined there continuously since that time. No intelligence test results available.

Case 21. Diagnosis: schizophrenia, paranoid type. A single male, thirty-five years of age, educated through high school and occupied as a pharmacy clerk. First psychotic symptoms in 1938, committed to Mt. Pleasant State Hospital in 1939. No intelligence test results available.

Case 22. Diagnosis: Schizophrenia, paranoid type. A married female, aged thirty-two years, educated through eleventh grade at seventeen, worked as a telephone operator until her marriage. First psychotic symptoms in 1938; spent two months in a private sanitarium early in 1939, and committed to Mt. Pleasant State Hospital in May, 1939. Scored Intelligence Quotient of 86 on Revised Stanford-Binet, Form L, vocabulary Average Adult. Classification: Dull Normal.

Case 23. Diagnosis: Schizophrenia, hebephrenic type. A single female, aged thirty-five years, educated through high school and spent several years in a convent; had also been occupied as a maid in a hospital. Admitted to Mt. Pleasant State Hospital for the first time in 1922 and discharged in 1932, re-admitted in 1936 and has remained there continuously since that time. No intelligence test results available.

Case 24. Diagnosis: Schizophrenia, hebephrenic type. A single female, forty-two years of age, educated through eighth grade. First psychotic symptoms in 1931, committed to Mt. Pleasant State Hospital in 1934. No intelligence test results available.

The individuals comprising the second group were freshmen students at the State University of Iowa selected on the basis of their scores on the Iowa Qualifying and Placement Examinations given in September, 1939.[6] They all ranked from the 90th to the 99th percentile on the Composite Score of the examinations, the percentiles being based on the scores made by the freshmen students taking the examinations that year. An unpublished study by Mitchell (17) indicated a correlation of .76 between the Intelligence Quotients of sixty-six freshmen, as scored on the Revised Stanford-Binet, Form L, and the Composite Scores on the Iowa Qualifying and Placement Examination, the average Intelligence Quotient being 122. The freshmen used in the present study may be regarded as generally comparable, although somewhat superior in terms of the test scores in question, to Mitchell's freshmen students.

Of the twenty-four freshmen, twelve were male and twelve were female; they ranged in age from seventeen years, five months to twenty-three years, one month. They came from homes in which the following occupations were represented by the wage-earners in the families: farmer, railroad engineer, jeweler, life insurance agent, plumber, piano tuner, attorney, professor of physiology, switchman, banker, shoe clerk, assistant postmaster, school teacher, real estate salesman,

[6] Three of the twenty-four freshman subjects were taken from the entering class of September, 1938, and one from the entering class of September, 1940. In each case the subject wrote while he was a freshman.

cashier of bank, electrician, and clerical worker.

Written language samples of 2800 words in length were obtained from all of the subjects in the following manner. These instructions were read to the subject: "You are to write a story of your life. Start at the beginning and write it just as you remember things. Any words will do. Even things that may seem unimportant to you should be written and especially things that have made a difference in your life. No one else will see what you have written." Then a copy of the instructions was given to the subject so that he could refer to them again. Each subject was told that his story should be at least 2800 words in length. When a subject did not write enough or asked further questions, instructions were continued in the above terms, or neutral comments were made. With most of the subjects more than one sitting was necessary in order for them to write samples of the length required.

In order to secure the written language samples, the patients were taken into a room off the ward in the hospital, and the freshmen were asked to come to a conference room in one of the university buildings. The writer secured the data from all subjects except the male patients, from whom the language samples were obtained by a male attendant in the hospital. Consistently undisturbing conditions were maintained insofar as was possible, and to a practically sufficient degree, while the samples were being written. Not more than six subjects were writing at the same time in a large-sized room, and the average total time required of the patients to write the sample of the required length was approximately eight hours, while the freshmen averaged approximately five hours. All subjects were cooperative for the most part, although the patients as a group were slower in beginning to write and less consistent in keeping at it, and therefore required more attention and encouragement. In no case, however, were topics suggested to the subjects or 'coaching' resorted to in order to obtain the requisite length of sample. The total time elapsed during the securing of the samples was approximately two weeks for the patients and approximately one month for the freshmen (with the exception of the four freshmen mentioned in the previous footnote).

The 2800-word samples were typed exactly as they were written. Each sample was then divided into twenty-eight successive one-hundred-word segments by counting the first one hundred words, placing a mark, and then counting the second hundred words, etc. Each word was then tabulated on sheets so designed that each one-hundred-word segment could be tabulated separately.[7] The procedure followed in tabulating the data was as follows.

After a sample had been typed, double-spaced, one of a pair of workers (much of the time one worker performed these tasks alone) placed numbers, one to one hundred, over the first one hundred words. These numbers, one over each word, were written very small. After the one hundredth word a small number one was written and encircled—to indicate the *limit* of the *first* one hundred words. The other worker, meantime, had written a letter of the alphabet in the upper-left hand corner of each of several tabulation sheets. The first word of the first one hundred words was noted and worker No. 1 looked all through the

[7] A copy of the tabulation sheet is in the appendix of the manuscript copy of this report which is on file in the State University of Iowa Library.

one-hundred-word sample, counting the number of times the word appeared. Worker No. 2 wrote this word, followed (in parenthesis) by the part of speech it represented in the "Word" column on the tabulation form that carried the letter of the alphabet under which the word would be classified alphabetically. The number of times the word appeared in the first hundred words was noted in the column headed "1" on the tabulation form. A small check was placed over the number, which had previously been placed over the word, as each word was counted.

After the first one hundred words had all been counted and tabulated, worker No. 1 counted off the next hundred words, numbering them from one to one hundred and placing an encircled 2 just after the last word of this second one-hundred-word section. Worker No. 2 totaled the frequencies noted in the column headed "1" on all of the tabulation forms used in order to check that the total was one hundred. The frequencies of the second one hundred words were noted in the column headed "2". Only the words appearing in this second one-hundred-word segment that did not appear in the first one hundred were written in the "word" column. This procedure was continued throughout the 2800-word sample.

The following rules were used in determining what constituted a word:

1. Each group of letters separated by spaces on both sides from adjacent groups of letters was counted as a word, even though it might be part of a place name, as in *Des* Moines (two words), an initial, as in James *A*. Brown (three words), or a neologism coined by a subject.
2. Any number was counted as one word; for example, 125 was tabulated as one word.
3. A hyphenated word was counted as one word, Webster's New International Unabridged Dictionary (23) being used as the authority as to whether or not a word should be hyphenated.
4. Each time a word was used as a different part of speech it was counted as a different word. For example, *mine* as a noun and *mine* as a pronoun were tabulated as two different words.
5. Common nouns and proper nouns having identical spellings were thrown together. For examples, the two words *Storm Lake* were tabulated under the common nouns *storm* and *lake*.
6. Contractions were divided into two words, for example, *didn't* was changed to *did not* and tabulated as two words.
7. Abbreviations which stood for only one word were written out and tabulated as the complete word. Abbreviations which consisted of more than one unit, as for example *M.D.* and *Ph.D.*, were tabulated as one word.
8. Misspellings, when it was apparent that they were misspellings and not neologisms were corrected and tabulated as corrected.

The part of speech was placed after a word as it was tabulated. Following is a list of the rules which were used in determining the part of speech represented by any given word. To be classified as:

Nouns—all regularly known common and proper nouns and gerunds which the dictionary[8] recognizes as nouns.

Pronouns—all personal and indefinite pronoun forms, including pronominal adjective forms, such as *my, our, your, their,* etc. Also all demonstrative, relative, and interrogative pronouns such as *this, those, who, whom, where,* etc.

Verbals—simple verbs, participles plus auxiliaries, gerunds and participles unless the dictionary recognizes them as nouns and adjectives, as the case may be.

Adjectives—regular classification, and any verb form (i.e. participle) which the dictionary recognizes as an adjective.

Adverbs—regular classification.

Prepositions—regular classification.

Conjunctions—regular classification.

Interjections—exclamatory expressions, and

[8] Webster's New International Unabridged Dictionary (23).

slang expressions used interjectionally.
Articles—*a, an* and *the.*

The data on the tabulation sheets were then analyzed and will be presented in three different sections: (1) Type-Token Ratios (TTR's), including both segmental TTR's and overall TTR's; (2) Grammatical Analysis; and (3) Type Frequencies.

The Type-Token Ratio

The type-token ratio[9] is a quantitative measure of language to which most attention has been given in the present study. The number of types in a given language sample is the number of *different* words occurring in the sample, and the number of tokens is the *total* number of words in the sample. The type-token ratio, then, is computed by dividing the number of different words by the total number of words in the sample. Since it may be assumed, from the work of Carroll (10), that the percentage of different words decreases as successive increments are added to a language sample, the number of tokens used in computing the type-token ratio must be kept constant in order to determine any variations within any given language sample, or in order to make the ratio comparable from one sample to another.

In this study the computations of the TTR's have been (1) the overall TTR as computed for the entire sample of 2800 words, and (2) the mean segmental TTR. As was stated previously, in this study each 2,800-word sample was divided into twenty-eight successive one-hundred-word segments. To secure the mean segmental TTR's the TTR was computed for each one-hundred-word segment independently and these segmental TTR's were averaged for each sample. This procedure makes it possible to compare samples of different magnitudes since such segmental TTR's are directly comparable as long as they represent segments of equal size, and the means of such segmental TTR's and mean segmental TTR's from the present study can be compared with those from any other study involving one-hundred-word segments, regardless of the number of such segments in a given sample.

Consideration of the TTR Scale

The limits of the TTR are mathematically defined as greater than zero and equal to or less than one. As to the nature of the cumulative TTR curve, it may safely be stated that D (the number of different words, or types) is a complex function of N (the total number of words, or tokens, in the sample). The greater the base on which the TTR is computed the smaller the absolute value of the TTR will be for any one sample of any one individual.[10]

The question may arise as to the relative value of the TTR unit at any given position on the scale from zero to one. This question is more obvious when it is considered whether the difference of one TTR unit at one point on the scale is equal to a difference of one TTR unit at any other point on the scale. First of all, the operational character of the TTR unit is clear. The question here raised would appear to be significant, if ever, whenever interpretations might be drawn as to the relation of the TTR to some other variable. It is to be pointed out that in any segment of the TTR scale where the variability of the TTR's for any given group of language samples is relatively large, a correspondingly larger absolute difference between any

[9] This term was introduced by Johnson (15) and the ratio has been discussed by him.

[10] The problems implied by these statements are treated in greater detail by Chotlos (11).

two TTR's would be required to satisfy the criteria of statistical significance, than would be required in any segment of the scale where the variability of TTR's is relatively less. In this sense, then, the question becomes one of the relative difference with regard to variability of TTR's at different points along the scale, and insofar as there are differences in such variability, it is to be expected that there will be corresponding differences in the relative significance statistically to be ascribed to differences of the same absolute magnitude, depending upon the segment of the scale which they involve. However, safeguards against misinterpretations that might conceivably result from this fact are to be found in the statistical procedures to be used in treating the data that are to be interpreted; it is not the similarity of two differences with regard to their absolute magnitude but the similarity with regard to their relative magnitude as shown in their degree of statistical significance, that would govern any interpretation regarding them. A logical consideration of the TTR scale would indicate that a mean TTR value at either the upper or lower end of the scale should imply a lower degree of variability among the individual TTR's of which it is the mean than would a mean TTR value in the middle range of the scale. This is true because variation from the mean in the direction of zero, in the case of a low mean TTR value, would obviously be limited in extent, and any relatively large variations from the mean in the opposite direction would tend to raise the mean; the same type of consideration would hold with regard to a relatively high mean TTR value. It is obvious, on the other hand, that a mean TTR value approximating .50, for example, does not necessarily imply any such limited range of deviations of the individual TTR's from the mean.

In order to make a partial investigation of the question under discussion Pearson product-moment correlations were run between the mean segmental TTR's and the standard deviations, separately for the psychotic subjects and for the freshmen. This correlation for the psychotic subjects was −.09, and for the freshmen it was −.12. Neither of these values deviates significantly from zero. This may be interpreted to mean that for each of the groups the TTR's fell within a segment of the scale within which there would appear to be no appreciable relation between the absolute magnitude of the TTR and its variability. However, the trend is in the direction indicated by the above logical considerations, and it may be assumed that the low correlations obtained are to be accounted for in part, at least, by the fact that TTR's for each group fell within a relatively narrow range.

In order to ascertain the degree of relation between the absolute value of the mean segmental TTR's and their variability when a larger number of measures and a larger range of the scale were involved, a Pearson product-moment correlation was run between the mean segmental TTR's and the standard deviations for all subjects. The correlation obtained was −.58. The fact that this correlation coefficient is higher than either of the corresponding coefficients for the separate groups tends further to substantiate the above logical considerations.

It is to be emphasized again, however, that the relationship implied by these coefficients of correlation and by the logical analysis of the scale are of no particular significance so far as the inter-

pretation of differences in TTR values is concerned, since the differences are to be interpreted with reference to their relative rather than their absolute magnitude. Misinterpretation would occur only if the indicated relationships were ignored; they are, of course, taken into account in the statistical procedures on the basis of which the significance of the differences between the TTR values is estimated.

Grammatical Analysis

The grammatical analysis is concerned with ascertaining the proportion of the entire language sample, for each subject and for each group of subjects, that is represented by each of the parts of speech. Relationships between certain parts of speech have been computed in terms of ratios.

Type Frequencies

The section on type frequencies is concerned with an objective language measure which expresses relative frequency of occurrence of each different word, or type. Of particular interest are those type frequencies which differentiate the written language of schizophrenic patients from that of freshmen. In order to select such types, if they exist, the one hundred most frequently used types were found for each group and comparisons of these were made. Particular attention was also given to certain types such as self-reference words and 'allness' terms, such as *never, always, all,* etc.

Also, the proportionate vocabularies of the two groups were compared. The proportionate vocabulary is found by determining the number of types making up a certain proportion of the tokens in a given language sample. Finally, a word list was compiled which presents each type separately and shows the number of subjects in each group who used the word, and the type frequency for each of the two groups.[11]

IV. RESULTS

Introduction to Results

In order to facilitate the discussion of the results the following system of symbols has been devised. The reader is asked to refer to this list for definitions of the symbols in terms of the operations to be performed in deriving the statistics which they represent.

The data were analyzed to determine the characteristics of the type-token ratios for one-hundred-word segments. The following symbols will be used in discussing the results of this section of the analysis.

Let $TTR = R = \frac{D}{N}$ where D is the number of different words (types) in a segment and N is the total number of words (tokens) in that segment.

Let R_p = segmental TTR where p is the subscript for any given one-hundred-word segment.

Let $R_1, R_2, R_3, \cdots R_p, \cdots R_{28}$ refer to segmental TTR's for each one-hundred-word segment, one through twenty-eight.

$$R_1 = \frac{D_1}{N_1},\quad R_2 = \frac{D_2}{N_2},\ \cdots\quad R_p = \frac{D_p}{N_p},\ \cdots\quad R_{28} = \frac{D_{28}}{N_{28}}$$

where $D_1, D_2, \cdots D_p, \cdots D_{28}$ are independently computed, the number of different words in any one segment not being influenced by any words in any other segment, and $N_1 = N_2 = \cdots N_p = \cdots N_{28} = 100$.

Let R_i represent the mean of $R_1, R_2, \cdots R_p \cdots R_{28}$ for each individual subject.

$$R_i = R_1 + R_2 + \cdots + R_p + \cdots + R_{28} = \frac{\Sigma R_p}{28}.$$

Let s_i represent the standard deviation of the $R_1, R_2, \cdots R_p, \cdots R_{28}$ for each individual subject.

$$s_i = \sqrt{\frac{\Sigma(R_p - R_i)^2}{28}}.$$

[11] This complete word list is contained in the appendix of the manuscript copy of this report on file in the State University of Iowa Library.

The data were analyzed in order to determine the characteristics of the segmental TTR for the group. The following symbols will be used in the discussion of the results of this part of the analysis.

Let R_m represent the segmental TTR for the group

$R_m = \frac{\Sigma R_i}{n}$ where R_i is the mean TTR per subject summed over all the group and n is the number of subjects in the group

Let s_m represent the standard deviation of the distribution of mean segmental TTR's (R_i's) for the group.

$$s_m = \sqrt{\frac{\Sigma(R_i - R_m)^2}{n}}.$$

Let $S.E._m$ represent the standard error of the group mean segmental TTR(R_m).

$$\text{Let } S.E._m = \sqrt{\frac{\Sigma(R_i - R_m)^2}{n(n-1)}} = \frac{\Sigma s_m}{\sqrt{n-1}}.$$

Let M_{si} represent the mean of the standard deviations for each of the n subjects in a group.

$$M_{si} = \frac{\Sigma s_i}{n}.$$

Let σ^2_{si} represent the estimated variance of the standard deviations for the group

$$\text{est. } \sigma^2_{si} = \frac{\Sigma(s_i - M_{si})}{n-1}.$$

The data were analyzed to determine the characteristics of the TTR when it is computed by considering the entire sample as a whole. This TTR is called the overall TTR and the following symbols will be used in discussing the results of this section of the analysis.

Let R' represent overall TTR

$R' = \frac{D'}{N'}$ where D' is the number of different words (types) and N' is the total number of words (tokens) in the entire sample. Computed independently for each subject.

Let R'_m represent the mean overall TRR for the group.

$R'_m = \frac{\Sigma R'}{n}$ where n is the number of subjects in the group.

Let s'_m represent the standard deviation of the overall TRR's for the group.

$$s'_m = \sqrt{\frac{\Sigma(R' - R'_m)^2}{n}}.$$

Let $\overline{S.E.}'_m$ represent the standard error of the group mean overall TRR.

$$\overline{S.E.}'_m = \sqrt{\frac{\Sigma(R' - R'_m)^2}{n(n-1)}}.$$

Let σ'^2_m represent the estimated variance of the overall TRR's for the group.

$$\sigma'^2_m = \frac{\Sigma R' - R'_m}{n-1}.$$

I. TYPE-TOKEN RATIO

Internal Consistency of Segmental TTR's

It was felt that it would be desirable to secure some measure or indication of the internal consistency (i.e. how well a random half of the sample measures what the whole sample measures) of the 2800-word samples for each subject. This was obtained by splitting the R_1, R_2, . . . R_p, . . . R_{28} for each subject at random into two sets of R's. The mean for each random half was computed and the *t*-test for related measures was applied.[12] It would be expected from such a test that if the internal consistency of the samples was high, the value of *t* so derived would fail to be statistically significant. When this test was applied to the random sets of R's for the patients the value of *t* was 1.82, and when applied to the random sets of R's for the freshmen

[12] See Lindquist, E. F., *Statistical Analysis in Educational Research*, Houghton Mifflin Company, Boston, 1940, p. 58. The procedure is that of finding the difference for each pair of R's and for this distribution of differences, determining whether or not the mean difference differs significantly from zero.

the value of t was .411. The values in both cases fall short of significance at the five per cent level of confidence with twenty-three degrees of freedom (d.f.).

Variability in Segmental TTR's

We were interested in determining whether the schizophrenic patients were more variable than the freshmen, not only from subject to subject, but whether they also showed more variability from segement to segment than did the freshmen. In order to determine this the s_1 for each subject and Ms_1 for each group were computed. The F ratio[13] when computed as a ratio of the variance of the distribution of s_1's for schizophrenic patients to the variance of the distribution of s_1's for freshmen resulted in a value of F of 10.35 which, with twenty-three and twenty-three d.f., is significant at the one per cent point.

TABLE 1
TTR's for each subject ranked in descending order within each group

Mean-Segmental TTR's		Overall TTR's	
Patients	Freshmen	Patients	Freshmen
.7450	.7357	.3932	.4079
.7404	.7354	.3854	.3907
.7386	.7339	.3618	.3607
.7164	.7307	.3596	.3471
.7007	.7293	.3407	.3457
.6975	.7286	.3154	.3454
.6846	.7279	.3150	.3450
.6757	.7261	.2961	.3439
.6700	.7236	.2946	.3411
.6700	.7236	.2821	.3375
.6700	.7200	.2789	.3307
.6668	.7196	.2779	.3293
.6657	.7143	.2746	.3289
.6618	.7118	.2725	.3250
.6607	.7104	.2639	.3229
.6582	.7082	.2575	.3218
.6482	.7057	.2464	.3104
.6436	.7054	.2371	.3089
.6389	.6975	.2371	.3086
.6264	.6946	.2279	.3014
.5993	.6943	.2121	.2971
.5700	.6932	.2121	.2921
.5346	.6836	.1943	.2879
.4600	.6708	.1850	.2689

In order to determine whether or not there was a difference in variability from segment to segment between the sexes, the F ratio was computed as a ratio of the variance of the distribution of s_1's for the female subjects to the variance of the distribution of s_1's for the male subjects. The value of F so obtained for the patients was 4.49 which with eleven and eleven d.f. is significant at the one per cent point, the male patients showing more variability than the female patients. The value of F so obtained for the freshmen was 1.47 which with eleven and eleven d.f. fails to be significant at the five per cent point, the value of F required for significance at that point being 2.82.

Means and Distributions of Mean Segmental TTR'S and Overall TTR's

The $R_1, R_2, \ldots R_p, \ldots R_{28}$ for each subject were averaged and a mean segmental R (R_1) obtained for each individual.[14] An overall R (R') was also obtained for each subject by considering the 2800-word sample as a unit and dividing the number of types in the entire sample by 2800. The R_1's and R''s for each group are ranked in descending order in Table 1. An examination of this table reveals that there is some overlapping between the two groups on both the R_1's and R''s. The R_1's for three patients were higher than the highest R_1 among the freshmen; and the R_1's for eight patients were higher than the lowest R_1 among the freshmen. The lowest R'

[13] See Lindquist, *op. cit.*, p. 60. F, the variance ratio, is defined as $\frac{\sigma_1^{12}}{\sigma_2^{12}}$ in which σ_1^{12} and σ_2^{12} are estimates of the true variances of the populations sampled.

[14] Table 1 in Appendix A of the manuscript copy of this report on file in the State University of Iowa Library presents the twenty-eight R's for each subject.

TABLE 2
Group means, standard error of means, and standard deviation for mean-segmental TTR's and overall TTR's

	Mean-Segmental TTR			Overall TRR		
	R_m	$S.E._m$	S_m	R'_m	$S.E.'_m$	S'_m
All Patients	.6559	.01322	.06404	.2801	.01180	.05625
Female	.6468	.02138	.06134	.2782	.01550	.05180
Male	.6651	.01608	.05385	.2819	.01855	.06078
All Freshmen	.7135	.00358	.01753	.3291	.00636	.03072
Female	.7179	.00392	.01254	.3350	.01060	.03548
Male	7091	.00590	.01776	.3232	.00710	.02365

among the freshmen was higher than the R''s for ten patients. Only one R' among the freshmen was higher than all R''s among the patients.

Table 2 presents the mean of the R_i's for the group (R_m) and the mean of the R''s for the group (R'_m) with the standard deviations (s_m for the R_i distribution and s'_m for the R' distribution) and the standard error of the means ($S.E._m$ for R_m, and $S.E.'_m$ for R'_m) for each group for all patients, female patients, male patients, all freshmen, female freshmen, and male freshmen.

The curves drawn from the frequency distributions of the twenty-four mean

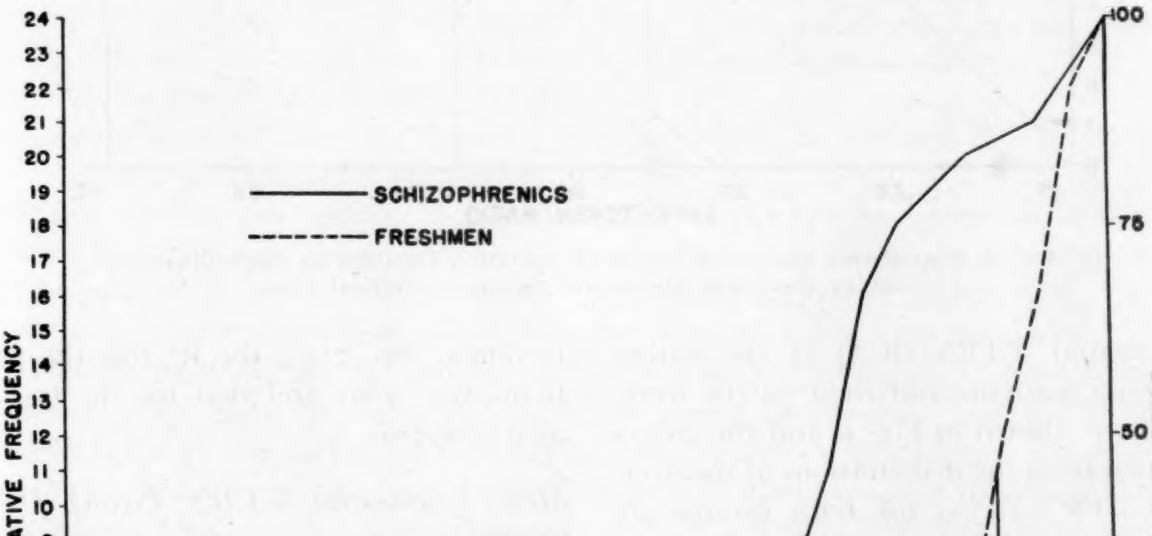


FIG. 1. Cumulative frequency curves of mean segmental TTR's for 24 schizophrenics and 24 freshmen. Means are shown by vertical lines

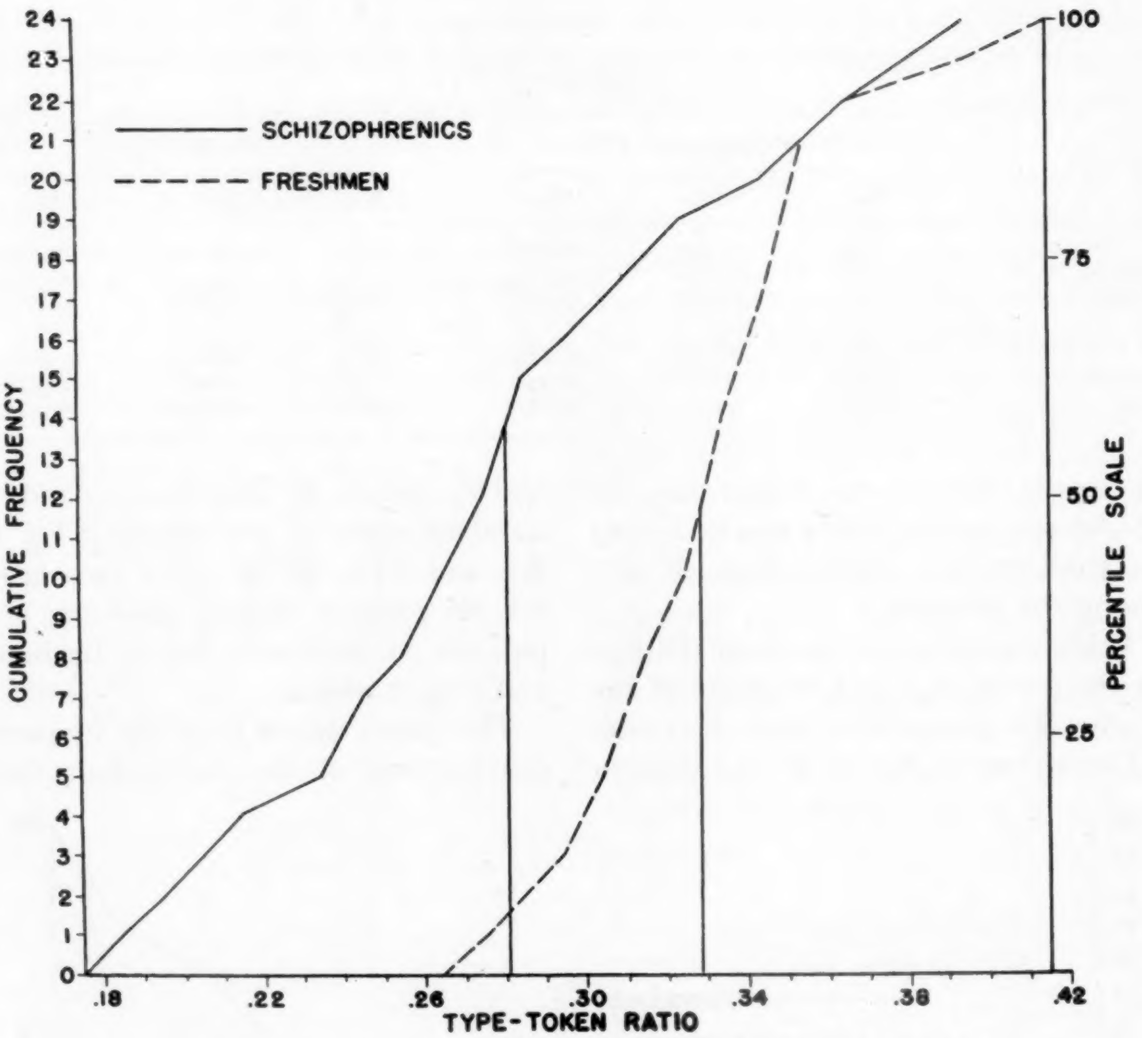


FIG. 2. Cumulative frequency curves of overall TTR's for 24 schizophrenics and 24 freshmen. Means are shown by vertical lines

segmental TTR's (R_i's) of the schizophrenic patients and those of the freshmen are shown in Fig. 1, and the curves drawn from the distributions of the overall TTR's (R′'s) for both groups are shown in Fig. 2. It is apparent from the curves in Fig. 1 that the range of R_i's for the patients is greater than that for the freshmen, the range for the patients being .4600 to .7450 while that for the freshmen is .6708 to .7357, indicating more variability among the patients. The range of the R′'s was also somewhat greater for the patients than for the freshmen, the values ranging from .1850 to .3932 for the patients, and from .2689 to .4079 for the freshmen. The R_m for the patients was .6559 while that for the freshmen was .7135; the R'_m for the patients was .2801 and that for the freshmen was .3291.

Mean Segmental TTR's: Group Differences

The *t*-test was applied to test the significance of the difference between R_m for the patients and R_m for the freshmen.[15] The value obtained for *t* was 4.204 which, with forty-six d.f., is significant at the one per cent level of confidence. Therefore, we would feel justified in rejecting the hypothesis that these samples were drawn from populations whose means are equal.

However, since one of the assumptions

[15] See Lindquist, *op. cit.*, p. 56-58.

underlying the *t*-test when used to test the significance of the difference between means of independent small samples is that the true variance of one sample must be equal (or approximately equal) to the true variance of the other sample, a test of the significance of the difference in variability was applied.

The F ratio when computed as the ratio of the variance of the distribution of R_i's for patients to the variance of the distribution of R_i's for the freshmen resulted in a value of F of 13.34. which with twenty-three and twenty-three d.f. is significant at the one per cent point. The results of this test indicate that the variability of the patients as a group exceeds the variability of the freshmen as a group to an extent which cannot be attributed to chance fluctuations in random sampling. Another way of stating this is that we are 'practically certain' that the samples are drawn from different populations and that our 'best estimate' of the true variance of the population from which the sample of schizophrenic patients was drawn is considerably greater than the corresponding 'best estimate' of the true variance of the population from which the sample of freshmen was drawn.

Although we have no way of knowing the *true* variance of the populations which have been sampled, there is some question as to the validity of applying *t* to test the significance of the difference in means in view of the difference in variability of the two groups. Therefore, the analysis of the data was extended to get a further indication of the significance of the difference between the means for the two groups which would not rest upon the assumption of homogeneity of variance. This was accomplished by using *t* to establish limiting values for each group outside of which any exact hypothesis as to the value of the true mean may be rejected with a given degree of confidence. At the one per cent level of confidence the limiting values of the true mean for the patients were .6188-.6930, and for the freshmen they were .7034-.7236. Since there is no overlap in these 'confidence intervals' we may be practically certain that the difference between the R_m's for the patients and for the freshmen indicates a real difference between the two groups.

The critical ratio of the difference between the R_m's for patients and for freshmen was 4.204. The probability that a C.R. of this magnitude for independently drawn samples from these two populations will be exceeded solely through errors in random sampling is .0001. This value of the critical ratio is larger than the criterion usually required for statistical significance.

The Effect of Certain Variables on TTR

In an attempt to determine how certain variables, particularly within the schizophrenic group, influence the TTR, the schizophrenics were sub-divided into groups and the average TTR's for these sub-groups compared. The fifteen patients for whom intelligence test results were available were split into two groups on the basis of I.Q., an I.Q. of 100 representing the dividing line. Seven patients had I.Q. scores below 100, ranging from 78 to 97, with an average of 87; eight patients had I.Q. scores above 100, ranging from 101 to 138, the average being 109. The mean segmental TTR's for the individuals within each group were averaged, resulting in an average mean segmental TTR of .6800 for the "above average I.Q." group and .6586 for the "below average I.Q." group. The *t*-test of the significance of this difference in the average mean segmental TTR's for

these groups resulted in a t value of .51 which with thirteen d.f. is not statistically significant.

Level of educational attainment would appear to be a variable among the patients which might influence the TTR's. To get an indication of how this factor affects the TTR, the patients were subdivided into three groups on the basis of level of educational attainment: ten patients who had college training; six patients who had graduated from high school but who had had no college training; and eight patients who had not graduated from high school. The average mean segmental TTR for each group was computed and a simple analysis of variance technique was used to determine whether the differences in means for the three groups are significant of real differences, or may be explained away in terms of chance fluctuations in random sampling. The mean segmental TTR's were .6462 for college graduates, .6876 for high school graduates, and .6395 for the lowest educational group, or non-high school graduates. The ratio (F) of the estimate of the populations variance, based on the variance of the group means, to the estimate of the population variance, based on variance within groups, resulted in an F value less than unity which obviously is not significant.

Duration of illness might conceivably be an important variable influencing the TTR's for the patients. Since duration of illness can at best be only roughly estimated, it was felt that the effect of duration of confinement in the hospital could be more reliably ascertained. Since the average length of confinement in the hospital was three years, the patients were sub-divided with this average as a criterion. The thirteen patients who had been confined in the hospital for a shorter period than three years had an average mean segmental TTR of .6579, while the patients who had been confined in the hospital three years or longer had an average TTR of .6536. The t-test of the significance of the difference between these means is not statistically significant.

Comparison of TTR's Computed from Written and Spoken Language

A study by Fairbanks (12) in which she compared mean segmental TTR's, using one hundred tokens as the size of the segment, for spoken language samples from schizophrenic patients and superior freshmen, yielded results comparable to those obtained in this investigation. These studies are highly similar in that the subjects used in both were drawn from the same populations and the procedures followed in tabulating and analyzing the data were essentially the same, but there is one point of difference which warrants some consideration. In her study Fairbanks employed an interview situation, involving the use of fourteen proverbs, the interview being recorded by means of an electric dictaphone technique without the subject's knowledge, and with instructions to the subjects to continue talking about anything that they wished to after finishing the proverbs. In the present study the instructions to the subjects were to write "the story of your life". It is doubtful that much importance should be attached to this difference in methods of securing the data, inasmuch as the samples of language obtained were probably of sufficient length to compensate for such differences.

In general, the main findings of Fairbanks as to the differences between the spoken language of schizophrenic patients and freshmen students were in the

same direction as those reported here for written language, the patients showing more variability as a group than the freshmen, and the difference between the mean segmental TTR's for the two groups being statistically significant.

Of particular interest is the comparison between spoken language and written language as indicated by these two studies. Table 3 presents the average mean segmental TTR's for each group and the range values, taken from the mean segmental TTR's for the individuals in each group, for written and spoken language of schizophrenic patients and freshman students. It may be readily observed from this table that the mean TTR for both types of subjects runs considerably higher for written language than for spoken language. This difference might have been anticipated because of the fact that in producing written material the individual has opportunity and ample time to alter and rearrange the words that he writes, which in many cases amounts to striving for variety or 'diversity' in the words used. Thus, this premeditated aspect of written language tends to obliterate the spontaneity which is more characteristic of spoken language.

It is interesting to note that the spoken language of freshmen is characterized by approximately the same mean segmental TTR value as is the written language of schizophrenics.

TABLE 3

The average mean segmental TTR's (R_m) for each group and the range values within each group for written and spoken language of schizophrenic patients and freshmen students. The data for spoken language are from Fairbanks (12)

	Written Language		Spoken Language	
	R_m	Range	R_m	Range
Schizophrenic Patients	.6559	.4600–.7450	.5681	.4933–.6193
Freshmen Students	.7135	.6708–.7357	.6416	.6137–.6650

Overall TTR's

In determining the differences in overall TTR's between the groups the steps in the analysis followed those presented for the differences in segmental TTR's for the groups.

The R'_m for the patients as a group was .2801 while that for the freshmen was .3291. The *t*-test applied to test the significance of the difference between these R'm's resulted in $t = 3.65$, which with forty-six d.f. is significant at the one per cent level of confidence. The results of this test indicate that the difference in R'_m's for the patients and for the freshmen is a real difference.

The F test of the significance of the difference between the variances of the distdribution of R″s for the two groups yielded a value of 3.35. The value required for significance at the one per cent point with twenty-three and twenty-three d.f. is 2.70. While the obtained value of F is greater than that required for significance at the one per cent point, it is not much greater.

The further test, which has been discussed previously, of using *t* to set limiting values of the true mean of each group was applied to the R'_m's of patients and of freshmen. The limiting values of the true mean for the patients at the one per cent level of confidence were .2370-.3132 while the limiting values of the true mean for the freshmen were .3113-.3469. There is a slight over-

lap in the intervals for the patients and the freshmen at the one per cent level of confidence, the upper limit for the patients extending .0019 above the lower limit for the freshmen. The limiting values of the true mean for the patients at the two per cent level of confidence were .2506-.3096 while those for the freshmen were .3132-.3450. Thus we are able to say that at the two per cent level of confidence there is a true difference between R'_m's for the patients and the freshmen.

The critical ratio of the difference between the R'_m's for patients and for freshmen resulted in a value of 3.654. The probability that a C.R. of this magnitude for independently drawn samples from these two populations will be exceeded solely through errors in random sampling is .0003. This test again indicates that the difference in R'_m's for the two groups is statistically significant.

Sex Differences

Since each group of twenty-four patients and twenty-four freshmen consisted of twelve male and twelve female subjects, the data were analyzed to determine whether there were significant differences between the sexes within each group for the R_m's and R'_m's.

The F ratio when computed as the ratio of the variance of the distribution of R_1's for the female patients to the variance of the distribution of R_1's for male patients gave an F value of 1.729 which, with eleven and eleven d.f., would be exceeded by chance in more than five per cent of similarly selected random samples. The results of this test give us no adequate basis for rejecting the hypothesis that these samples were drawn from equally variable populations. Likewise, the t-test of the significance of the difference between the R_m's for the female patients and the male patients resulted in a value of t of .683 which, with twenty-two d.f., is clearly not significant, since a value of this magnitude can be expected to occure by chance more than fifty per cent of the time in similarly selected random samples.

The F ratio when computed as the ratio of the variance of the distribution of R_1's for female freshmen to the variance of the distribution of R_1's for male freshmen gave an F value of 2.661 which, with eleven and eleven d.f., is not statistically significant. This value can be accounted for by chance fluctuations in random sampling and we are therefore not justified in rejecting the hypothesis that the samples were drawn from equally variable populations. Similarly, the t-test of the significance of the difference between R_m's for female freshmen and male freshmen resulted in a $t = 1.24$ which, with twenty-two d.f., is not statistically significant since a value of t of this magnitude can be expected to occur by chance between twenty and thirty per cent of the time. The results of this analysis give us no adequate basis for assuming any difference in R_m's between male and female patients or between male and female freshmen.

The ratio F when computed as a ratio of the variance of the distribution of R''s for female patients and the variance of the distribution of R''s for male patients resulted in an F of 1.431 which, with eleven and eleven d.f., would occur by chance more than five per cent of the time in similarly selected random samples. The t-test when applied to the difference in R'_m's for female and male patients resulted in $t = .153$ which, with the twenty-two d.f., would occur by chance more than eighty per cent of the time. The same tests when applied to the distributions of R''s for female and male

freshmen resulted in values of F = 2.231 which, with eleven and eleven d.f., is not statistically significant, and when applied to the difference in R'_m's for female and male freshmen the resulting $t = .922$ which, with twenty-two d.f., would be expected to occur by chance between thirty and forty per cent of the time in such samples. Our conclusion again is that the differences in R′ 's between the sexes for the two groups are not statistically significant and we are not justified in rejecting the hypothesis that the samples consisting of females and males, respectively, in each group were drawn from the same populations.

Correlation Between Mean Segmental and Overall TTR's

The Pearson product-moment correlation coefficient between the R_m and R'_m was .62 for patients and .62 for freshmen. For all forty-eight subjects $r = .71$. The fact that the r for all subjects is greater than the r for either the patient or the freshmen group may be due to the bimodality of the distribution for all subjects, or to the discrepancy in variability between patients and freshmen, or to an interaction of these two factors.

2. Grammatical Analysis

Distributions and Group Differences

The data were analyzed to determine the relative frequency of usage of each of the eight conventional parts of speech, plus articles (which were treated separately from other adjectives). Table 4 presents these frequencies, expressed as percentages of the total number of words (67,200) used by each group, separately for schizophrenic patients and for freshmen. The standard deviations of the distributions of the five main categories, and range values for each category taken from individual samples are also included in the table.

The statistical significance of the difference between the groups was tested for adjectives, adverbs, nouns, pronouns, and verbs. Of particular interest were the differences in percentages of the total number of words represented by each of these grammatical categories, and differences in variability of usage of these parts of speech.

The t-test was applied to test the significance of the difference between patients and freshmen in percentages for certain parts of speech. The values of t

Table 4

Relative frequency of usage of different parts of speech expressed as percentage of the total number of words used by the group (67,200), with standard deviations of the distributions of five main categories. The range values are from individual samples*

	Schizophrenic Patients			Freshmen Students		
	Percentages	S.D.	Range Values	Percentages	S.D.	Range Values
Nouns	24.27	3.98	17.43–33.68	22.15	2.26	17.86–25.57
Pronouns	13.12	3.78	4.79–20.25	14.57	1.50	11.68–17.07
Verbs	19.82	2.30	15.86–23.93	18.71	1.60	16.18–22.36
Adverbs	7.70	1.71	3.68–10.57	8.34	1.05	6.00–10.79
Adjectives	8.33	2.56	4.68–16.00	9.45	1.17	6.89–10.96
Conjunctions	7.23		3.75– 9.46	6.55		4.32– 8.29
Prepositions	12.33		7.75–16.57	12.35		10.46–14.43
Interjections	0.07		0.04– 0.86	0.05		0.00– 0.21
Articles	7.15		4.96–11.00	7.83		5.21–10.11

* Table 1 in Appendix B of the manuscript copy of this report on file at the State University of Iowa Library contains the percentage of usage of each part of speech for each individual.

TABLE 5

Values of t and F obtained from testing significance of the difference in usage of certain grammatical categories based on percentage of total sample between schizophrenic patients and freshmen

	Values of t	Values of F
Adjectives	1.88	4.78
Adverbs	1.54	2.68
Nouns	2.22	3.11
Pronouns	1.77	6.38
Verbs	1.86	2.07

$n=24$

With forty-six d.f. the values of t required for significance are: at the one per cent level of confidence $t=2.69$; at the five per cent level of confidence $t=2.01$. With twenty-three and twenty-three d.f. the values of F required for significance are: at the one per cent point $F=2.72$; at the five per cent point $F=2.01$

obtained for the categories tested are presented in Table 5. The only t value which might possibly be regarded as statistically significant is that obtained for the difference in percentage of nouns. This t is significant at the five per cent level of confidence. The differences between schizophrenic patients and freshmen in percentages for adjectives, adverbs, pronouns, and verbs may, therefore, be attributed to chance fluctuations in random sampling.

The F ratio when computed as the ratio of the variance of the distribution of percentages (based on total words per sample) for each grammatical category, used by the patients, to the variance of the distribution of percentages of the same category used by the freshmen, resulted in the values of F presented in Table 5. Each F value was statistically significant, the F values obtained for adjectives, nouns, and pronouns being significant at the one per cent point while those for adverbs and verbs were significant at the five per cent point. We may conclude that the variability of the patients as a group exceeded that of the freshmen as a group in relative frequency of usage of five grammatical categories, by an amount which cannot be attributed to chance fluctuations in random sampling.

Sex Differences

Table 6 presents the relative frequency of usage of different parts of speech expressed as percentage of the total number of words used by each sex (33,600) in each group, with the standard deviations of the distributions of the five main categories.

The t-test was used to test the significance of the differences between males and females in each group in relative

TABLE 6

Relative frequency of usage of different parts of speech expressed as percentage of the total number of words used by each sex (33,600), with standard deviations of the five main categories

	Schizophrenic Patients				Freshmen Students			
	Female		Male		Female		Male	
	Percentages	S.D.	Percentages	S.D.	Percentages	S.D.	Percentages	S.D.
Nouns	23.73	4.61	24.80	3.20	22.77	2.18	21.53	2.15
Pronouns	13.99	4.39	12.26	2.85	14.27	1.71	14.87	1.81
Verbs	20.22	2.59	19.39	1.99	18.47	1.80	18.95	1.33
Adverbs	7.93	1.87	7.47	1.53	8.09	0.76	8.59	1.30
Adjectives	8.43	2.02	8.24	3.03	9.36	0.86	9.53	1.48
Conjunctions	6.88		7.58		6.99		6.12	
Prepositions	12.00		12.65		12.26		12.44	
Interjections	0.09		0.04		0.05		0.05	
Articles	6.73		7.57		7.74		7.92	

TABLE 7

Values of t and F obtained from testing significance of the difference, in usage of grammatical categories based on percentage of total sample between the sexes within each group

	Schizophrenic Patients		Freshmen Students	
	Values of t	Values of F	Values of t	Values of F
Adjectives	.174	2.24 (Males)*	.333	2.93 (Males)
Adverbs	.638	1.53 (Females)	1.088	2.91 (Males)
Nouns	.633	2.07 (Females)	1.333	1.03 (Females)
Pronouns	1.094	2.37 (Females)	.907	1.73 (Females)
Verbs	.846	1.69 (Females)	.706	1.85 (Females)

With twenty-two d.f. the values of t required for significance are: at the one per cent level of confidence $t=2.819$; at the five per cent level of confidence $t=2.074$.

With eleven and eleven d.f. the values of F required for significance are: at the one per cent point $F=4.46$; at the five per cent point $F=2.82$.

* The sex which was more variable in each case.

frequency of usage of certain parts of speech. The values of t, presented in Table 7, are not statistically significant for any of the grammatical categories tested either for schizophrenic patients or for freshmen.

The F ratio when computed as the ratio of the variance of the distribution of percentages for each grammatical category (based on total words per sample) for female subjects, to the variance of the distribution of percentages for the same category for male subjects, resulted in values shown in Table 7, for patients and for freshmen. The F values obtained for adjectives and adverbs as between male and female freshmen exceed the value of F required for significance at the five per cent point. In each of these two categories the male freshmen were

TABLE 8

Comparison of the relative frequency of usage of parts of speech in written and in spoken language expressed as percentage of the total number of words used by the groups, 67,200 in the case of written and 30,000 in the case of spoken language. Data for spoken language from Fairbanks (12)

	Spoken		Written	
	%	Range	%	Range
Schizophrenic Patients				
Nouns	13.04	10.40–16.63	24.27	17.43–33.68
Pronouns	22.68	19.33–24.75	13.12	4.79–20.25
Verbs	26.28	24.27–30.47	19.81	15.86–23.93
Adverbs	11.54	7.00–17.97	7.70	3.68–10.57
Adjectives	5.37	3.77– 7.10	8.33	4.68–16.00
Conjunctions	6.55	4.10– 8.77	7.23	3.75– 9.46
Prepositions	7.48	4.30–10.00	12.33	7.75–16.57
Interjections	2.64	0.53– 4.43	0.07	0.04– 0.86
Articles	4.48	2.53– 6.87	7.15	4.96–11.00
Freshman Students				
Nouns	15.39	12.67–18.53	22.15	17.86–25.57
Pronouns	17.96	14.40–20.40	14.57	11.68–17.07
Verbs	22.95	20.50–24.47	18.71	16.18–22.36
Adverbs	10.16	8.87–11.20	8.34	6.00–10.79
Adjectives	6.69	5.67– 7.87	9.45	6.89–10.96
Conjunctions	8.83	7.33–11.40	6.55	4.32– 8.29
Prepositions	10.00	8.80–11.00	12.35	10.46–14.43
Interjections	1.26	0.47– 2.00	0.05	0.00– 0.21
Articles	6.79	5.27– 9.07	7.83	5.21–10.11

TABLE 9

Rank order of increase in relative frequency of usage of parts of speech, expressed as percentage of the total number of words used by the group in written over spoken and spoken over written language. Schizophrenic patients and freshmen students. Data for spoken language from Fairbanks (12).

Rank Order of Increase (Written over Spoken)			
Schizophrenic Patients	%	Freshman Students	%
Nouns	86.1	Nouns	43.9
Prepositions	64.8	Adjectives	41.3
Adjectives	55.1	Prepositions	23.5
Articles	59.6	Articles	15.3
Conjunctions	10.4		
Rank Order of Increase (Spoken over Written)			
Schizophrenic Patients	%	Freshman Students	%
Pronouns	72.9	Verbs	27.7
Verbs	32.7	Pronouns	23.3
Adverbs	49.8	Conjunctions	34.8
Interjections	3671.4	Adverbs	21.8
		Interjections	2420.0

more variable than the female freshmen.

Comparison of Written and Spoken Language

Table 8 presents a comparison of the relative frequency of usage of parts of speech in written language with that in spoken language, the latter data being taken from the above mentioned study by Fairbanks (12) concerned with the spoken language of schizophrenic patients and freshman students. This comparison is justified by the fact that the data presented from Fairbanks' study were from samples drawn from the same two general types of subjects and were analyzed in essentially the same manner as were the data presented in this study. This latter consideration is of great importance in view of the fact that results in word count studies and grammatical usage analyses depend to a large extent upon the rules followed in determining what constitutes a word and the rules used in classifying words as to the parts of speech represented by them. An examination of Table 8 reveals several differences in the relative frequency of usage of parts of speech in the spoken and written language of schizophrenic patients and freshman students, respectively. These differences are summarized in Table 9 by showing the rank order of increase in usage of the various parts of speech in written over spoken and spoken over written language for each of the two groups.

There is a marked increase in percentage of nouns, adjectives, prepositions, and articles for both schizophrenics and freshmen, and in conjunctions for schizophrenics, in written language over spoken language. For both groups the amount of increase in written over spoken language is greatest for the nouns, the patients showing 86.1 per cent increase and the freshmen 43.9 per cent increase in nouns used in written over spoken language.

There is an increase in percentage of pronouns, verbs, adverbs, and interjections for both groups, and in conjunctions for the freshmen, in spoken over written language. The largest amount of increase in spoken over written language was 72.9 per cent in the pronouns for the patients and 27.7 per cent increase in verbs for the freshmen. (The increase for interjections, for both groups, was so great as to mean for all practical purposes that interjections are used only in spoken language.) The parts of speech for which there was increase in written over spoken language and increase in spoken over written language were the same for the two groups wtih the exception of conjunctions, which showed a slight increase in written over spoken language for the patients, and an increase in spoken over written language for the freshmen.

Inter-relationships Among Parts of Speech

Of the relationships between certain parts of speech, the adjective-verb quotient (A_{vq}) is of perhaps the greatest interest, since it, or a variation of it, has been used by other investigators. Busemann (4), as reported by Boder (3), recorded in shorthand a number of stories told by children of different ages and found a marked fluctuation of the relationship between 'qualitative' and 'active' (dynamic) expressions. In the category of qualitative expressions he included not only adjectives, but also nouns and participles of verbs, when used as attributes to any other nouns; in the category of active expressions he included all verbs except the auxiliary. By dividing the number of verbs by the number of qualitative expressions he obtained a measure which he called the Action quotient (Aq.) of style. Busemann found that a rhythmical increase and decrease of the Aq. occurs with increase in age, which he believes to correspond to alleged rhythmical changes of emotional stability during childhood, adolescence, and youth. Furthermore, according to Busemann's theory, these rhythmical variations continue throughout the whole lifetime and reflect rhythmical variations of emotional stability and creative power.

Rorschach (19), again as reported by Boder (3), in classifying the interpretations given by subjects to a series of ink blots, calculated the ratio between different types of descriptions made. He found that the predominance of kinaesthetic description (verbs) indicates moderate, sluggish motility, introversion, and little adaptability to reality, while the predominance of color descriptions (qualitatives) reflects the excited, but alert, exact, and rapid motility, extraversion, and better adjustment to reality.

Stimulated by the suggestions made in these studies, Boder (3) set out to find whether there exist gross differences of adjective-verb ratios corresponding to differences in subject matter of various classes of writing. He inverted the procedure of Busemann, however, and took the adjective as the numerator in order to obtain a measure which might (if Busemann is right) correlate positively with desirable traits. The ratio he used indicates the number of adjectives per one hundred verbs and is designated in purely grammatical (as opposed to Busemann's behavioral 'action quotient') terms as the Adjective-Verb Quotient (A_{vq}). He found that for each of the kinds of writings studied, i.e., plays, legal statutes, fiction, and scientific monographs, the distribution of $A_{vq.}$'s shows sufficiently large differences to prove that as a rule the $A_{vq.}$ varies with the subject matter of the text. The Adjective-Verb Quotients reported in the present study are fairly comparable to the quotients reported by Boder, although the special rules followed by him in the word count analyses were somewhat different from the ones followed in the present analysis. The main differences were that in his study only attributive adjectives were counted, i.e., only adjectives placed before the noun; quantitative and ordinal numerals were not counted; no forms of *have* and *be* were counted, nor were *could, should,* and *would*. Inasmuch as the rules followed in the present study differed from those of Boder in such a way as to increase both the number of adjectives and the number of verbs, we might expect the ratios to remain fairly comparable as between the two studies.

Table 10 presents the A_{vq}'s for all scizophrenic patients and all freshman students ranked in descending order.

TABLE 10

Adjective-verb quotients for schizophrenic patients and freshman students for written language, ranked in descending order for each group

Adjective-Verb Quotients: Schizophrenics	Freshmen
.93	.66
.92	.64
.58	.62
.53	.62
.53	.60
.51	.59
.51	.58
.49	.58
.42	.55
.42	.54
.41	.54
.39	.53
.39	.53
.36	.52
.35	.50
.34	.48
.33	.48
.31	.47
.30	.42
.30	.41
.30	.37
.29	.35
.29	.35
.22	.35

With the exception of two patients whose $A_{vq.}$'s were strikingly high, the $A_{vq.}$'s for six freshmen were higher than those of the patients, and the $A_{vq.}$'s for nine patients were lower than the lowest one for the freshmen. Table 11 presents the mean $A_{vq.}$'s for both groups for written and spoken language, together with the mean quotients for adjectives to nouns, and adverbs to verbs for both groups for written and spoken language. Although the values for both of the latter quotients were larger for the freshmen than for the patients, indicating the use of more adjectives per noun, and more adverbs per verb, these quotients did not appear to be as differentiating as between freshmen students and schizophrenic patients as did the adjective-verb quotients.

The *t*-test was used to test the significance of the difference in mean $A_{vq.}$'s derived from written language for patients and freshmen, resulting in a value of 1.93 which, with forty-six d.f., is almost significant at the five per cent level, the value needed for significance being 1.95.

Table 12 shows the comparison of the mean $A_{vq.}$'s for schizophrenic patients and freshmen students for both written and spoken language, together with the average $A_{vq.}$'s obtained by Boder for each of four different types of style of writing. This table reveals that the mean $A_{vq.}$ for the spoken language of schizophrenic patients falls slightly below that of Boder's 'normative' style, while the mean $A_{vq.}$ for freshman students on spoken

TABLE 11

Relationships between certain parts of speech expressed as ratios for each group. The ratios for spoken language were computed from Fairbanks' data (12)

	Written: Schizophrenic Patients	Written: Freshmen Students	Spoken: Schizophrenic Patients	Spoken: Freshmen Students
$\frac{\text{Adjective}}{\text{Verb}}$ quotient	.43	.51	.20	.29
$\frac{\text{Adjective}}{\text{Noun}}$ quotient	.34	.42	.41	.43
$\frac{\text{Adverb}}{\text{Verb}}$ quotient	.39	.44	.44	.44

TABLE 12

Comparison of the A_{vq}'s obtained from written and spoken language of schizophrenic patients and freshmen students with those obtained by Boder

	Obtained Values of A_{vq}.
Schizophrenics, Written	.43
Freshmen, Written	.51
Schizophrenics, Spoken	.20
Freshmen, Spoken	.29
Boder's Data:	
Conversational (drama)	.11
Normative (legal statutes)	.20
Narrative (fiction)	.35
Descriptive (science)	.76

material falls midway between Boder's 'normative' and 'narrative' styles. The A_{vq}.'s computed from written language samples are considerably higher than those computed from spoken language for both schizophrenic patients and for freshman students. The mean A_{vq}. for schizophrenic patients on written material falls somewhat above that for Boder's 'narrative' type, while the mean A_{vq}. for freshmen on written material falls about midway between the A_{vq}.'s for Boder's 'narrative' and 'descriptive' types. The differences between written as opposed to spoken language for both groups correspond to the findings of Boder. He suggests that this may be explained by the fact that

"the time of writing is under the author's control; so that he can pay more attention to the style and choose the proper expressions. He has the possibility of rereading his material and inserting adjectives where found necessary, thus converting his material into a product of repeated and premeditated activity, lacking the spontaneity and speed which characterize the dialogue."(3)

3. TYPE FREQUENCIES

Table 13 presents a list of the hundred most frequently used words for the schizophrenic patients and the freshmen students, respectively. The list for the freshmen has those words common to both lists arranged in order of frequency, while the list for the schizophrenics has the words corresponding to those of the freshmen arranged in order of sequence regardless of frequency. The seventeen words in each of the two groups not common to both lists are arranged at the bottom of the table in order of frequency. When this list of one hundred most frequently used words in written language of these two groups is compared with that reported by Fairbanks (12) for spoken language, we find that sixty-nine of the hundred are common to both lists for freshmen and sixty-four of the hundred are common to both lists for schizophrenic patients.

Fairbanks reported some striking differences in the frequencies with which certain types occurred in the spoken language of schizophrenics and freshmen. She found, for example, that schizophrenics used *not* almost twice as many times as did the freshmen, and that *no* and *never* occurred in the schizophrenic list while *not* was the only negative word that occurred among the one hundred words most frequently used by the freshmen. An examination of the words in Table 13 shows that these group differences are not found in the written language. Fairbanks also reported that *very* was used three times more often by freshmen than by schizophrenic patients, while in the present study the patients used *very* almost three times more often than did the freshmen.

Table 14 shows the relative frequency of occurrence of first person singular pronouns (*I, my, mine, me, myself*), first person plural pronouns (*we, our, ours, us, ourselves*), second person pronouns, singular and plural (*you, your, yours, yourself, thee, thou*), and third person pronouns, singular and plural (*he, his,*

TABLE 13

List of 100 words most frequently used by schizophrenics and freshmen. The first 83 words common to both lists are arranged in descending rank order according to frequency of usage for freshmen. The remaining 17 words not common to both lists are arranged in order of frequency for the two groups at the end of the Table

	Freshmen			Schizophrenics		
	Word	Part of Speech	Freq.	Word	Part of Speech	Freq.
1.	the	art.	3354	the	art.	3052
2.	I	pro.	2778	I	pro.	2662
3.	and	conj.	2350	and	conj.	2950
4.	to	prep.	1805	to	prep.	2093
5.	was	verb	1468	was	verb	1069
6.	my	pro.	1346	my	pro.	859
7.	in	prep.	1328	in	prep.	1054
8.	of	prep.	1162	of	prep.	1641
9.	a	art.	844	a	art.	847
10.	it	pro.	672	it	pro.	507
11.	we	pro.	646	we	pro.	795
12.	had	verb	603	had	verb	646
13.	not	adv.	552	not	adv.	468
14.	that	pro.	442	that	pro.	430
15.	with	prep.	440	with	pro.	457
16.	at	prep.	429	at	prep.	417
17.	for	prep.	428	for	prep.	139
18.	have	verb	421	have	verb	416
19.	on	prep.	400	on	prep.	326
20.	were	verb	399	were	verb	278
21.	but	conj.	395	but	conj.	215
22.	that	conj.	387	that	conj.	170
23.	school (s)	noun	371	school (s)	noun	288
24.	this	pro.	327	this	pro.	236
25.	time	noun	289	time	noun	287
26.	is	verb	269	is	verb	585
27.	which	pro.	269	which	pro.	173
28.	when	conj.	252	when	conj.	297
29.	would	verb	245	would	verb	350
30.	she	pro.	239	she	pro.	180
31.	our	pro.	228	our	pro.	150
32.	did	verb	221	did	verb	184
33.	an	art.	212	an	art.	158
34.	from	prep.	211	from	prep.	264
35.	he	pro.	210	he	pro.	298
36.	her	pro.	198	her	pro.	144
37.	by	prep.	196	by	prep.	192
38.	or	conj.	192	or	conj.	241
39.	year	noun	189	year	noun	111
40.	mother	noun	182	mother	noun	107
41.	as	conj.	181	as	conj.	319
42.	they	pro.	181	they	pro.	243
43.	first	adj.	167	first	adj.	102
44.	one	adj.	165	one	adj.	174
45.	us	pro.	163	us	pro.	122
46.	do	verb	162	do	verb	210
47.	about	prep.	159	about	prep.	169
48.	them	pro.	159	them	pro.	163
49.	so	adv.	156	so	adv.	119
50.	out	adv.	155	out	adv.	194
51.	who	pro.	151	who	pro.	86
52.	his	pro.	149	his	pro.	132
53.	as	adv.	142	as	adv.	153
54.	all	noun	140	all	noun	87
55.	one	pro.	139	one	pro.	111
56.	life	noun	134	life	noun	123
57.	years	noun	134	years	noun	130

TABLE 13 (Continued)

	Freshmen			Schizophrenics		
	Word	Part of Speech	Freq.	Word	Part of Speech	Freq.
58.	two	adj.	131	two	adj.	138
59.	then	adv.	130	then	adv.	186
60.	up	adv.	129	up	adv.	99
61.	could	verb	128	could	verb	111
62.	went	verb	124	went	verb	290
63.	remember	verb	120	remember	verb	118
64.	too	adv.	114	too	adv.	178
65.	other	adj.	113	other	adj.	88
66.	are	verb	112	are	verb	224
67.	some	adj.	112	some	adj.	137
68.	all	adv.	111	all	adv.	87
69.	always	adv.	110	always	adv.	94
70.	good	adj.	103	good	adj.	152
71.	am	verb	100	am	verb	95
72.	so	conj.	100	so	conj.	139
73.	after	prep.	97	after	prep.	110
74.	there	adv.	95	there	adv.	353
75.	day (D)	noun	96	day (D)	noun	126
76.	very	adv.	93	very	adv.	252
77.	what	pro.	93	what	pro.	96
78.	if	conj.	92	if	conj.	109
79.	just	adv.	92	just	adv.	86
80.	go	verb	91	go	verb	159
81.	took	verb	90	took	verb	101
82.	quite	adv.	86	quite	adv.	90
83.	get	verb	85	get	verb	102
84.	home	noun	172	house	noun	181
85.	high (H)	adj.	169	young	adj.	152
86.	little	adj.	104	got	verb	150
87.	never	adv.	97	work	noun	134
88.	teacher	noun	97	also	adv.	128
89.	into	prep.	93	used	verb	126
90.	more	adv.	92	Iowa	noun	125
91.	any	adj.	90	people	noun	120
92.	class	noun	90	can	verb	113
93.	things	noun	90	been	verb	112
94.	only	adv.	85	father	noun	105
95.	still	adv.	85	him	pro.	99
96.	came	verb	84	city (C)	noun	90
97.	made	verb	82	will	verb	93
98.	than	conj.	81	while	conj.	82
99.	town	noun	81	know	verb	82
100.	great	adj.	80	like	verb	81

TABLE 14

Relative frequency of usage of the different personal pronouns in spoken and written language expressed as percentage of the total number of words used by each group, 30,000 for each group on spoken material, and 67,200 for each group on written material. Data for spoken language from Fairbanks (12)

	Spoken		Written	
	Schizophrenics	Freshmen	Schizophrenics	Freshmen
	%	%	%	%
First person singular	10.42	3.69	5.92	7.05
First person plural	.34	1.05	1.59	1.57
Second person singular and plural	1.44	2.14	.16	.06
Third person singular and plural	5.52	6.41	2.81	3.13

him, himself, she, her, hers, herself, it, its, itself, they, their, theirs, them) in the written language of schizophrenic patients and freshman students and Fairbanks' spoken language. It is apparent from this tabulation of the data that her findings in regard to differences between the groups in spoken language were not substantiated with regard to written language.

Proportionate Vocabulary

From her spoken language data Fairbanks found that the schizophrenics used only thirty-three types to make up fifty per cent of the total number of tokens, while the freshman group used forty-six types to arrive at the same percentage. In the present study of written language, the schizophrenics used ninety-five types to make up fifty per cent of the total number of tokens, while the freshman group used ninety-six types to make up the same percentage. For both groups ten types make up slightly over twenty-five per cent of the tokens in the written language. In connection with these comparisons of proportionate vocabulary of written and spoken language it should be pointed out that the number of tokens used by each group for the written language data was 67,200 while for the spoken language data the number of tokens for the schizophrenics was 29,800 and for the freshmen it was 30,000.

By dividing the number of types making up fifty per cent of tokens by the total number of tokens in each case the following percentages were obtained: for written language, .14 for both freshmen and for patients, and for spoken language, .15 and .11 for freshmen and for schizophrenics, respectively.

The patients used fifty-seven words which appeared to be privately coined words or neologisms while the freshmen used only five words which might be considered neologisms.[16]

V. SUMMARY AND CONCLUSIONS

This study is concerned primarily with the specific problem of determining whether and in what respects 'adequate' and 'inadequate' language might be differentiated quantitatively. Twenty-four schizophrenic patients, twelve male and twelve female, were selected to represent a group presenting 'inadequate' language and twenty-four superior university freshmen, twelve male and twelve female, were selected to represent a group presenting relatively 'adequate' language.

A 2800-word written language sample was obtained from each of the subjects under as uniform conditions as possible, the instructions to the subjects being to "write a story of your life." Each sample thus obtained was divided into twenty-eight successive one-hundred-word segments and each word, together with the part of speech it represented, was tabulated on sheets so designed that each one-hundred-word segment was recorded separately. Three types of analysis of the data were made: (1) the type-token ratio which is computed by dividing the number of different words (types) by the total number of words (tokens) in a given sample. In this study the ratio was computed for each one-hundred-word segment and the twenty-eight segmental

[16] Table 1 in Appendix C of the manuscript copy of this report on file at the State University of Iowa Library contains an alphabetical word list showing the number of freshmen and/or schizophrenics who used each word, and the frequency of its occurrence in each group. Words starred in the list are words which were considered neologisms in the generally used sense of that term; that is, they were privately coined by the individuals who used them and are not used by other persons. The starred words of the freshmen are mainly slang terms essentially, although relatively unusual.

TTR's obtained from each sample were averaged to secure a mean segmental TTR for each individual. An overall TTR was also obtained for each individual by considering the 2800-word sample as a unit and dividing the number of different types in the entire sample by 2800. (2) Grammatical Analysis; and (3) Type Frequencies. Statistical treatment of the data resulted in the following findings.

Type-Token Ratios

1. When the twenty-eight segmental TTR's for each subject were split at random into two sets of TTR's, the mean for each random half computed, and the *t*-test for related measures applied, it was found that the difference between the mean segmental TTR's yielded by the two random sets was not statistically significant for the patients nor for the freshmen.

2. The standard deviation of the twenty-eight segmental TTR's for each subject was computed. When the F-test of the significance of the difference in variability was applied it was found that the schizophrenic patients showed significantly more variability in the number of types used per one-hundred-word segment than did the freshmen.

3. When the mean segmental TTR and the overall TTR for each subject were compared it was found that the overall TTR's were consistently lower for all subjects than the mean segmental TTR's, bearing out the assumption that as an individual's verbal output increases the rate of increase in the number of different words he uses tends to decrease. There was some overlapping between the schizophrenic patients and freshmen on both the mean segmental TTR's and the overall TTR's, the range of values for mean segmental TTR's being .4600 to .7450, and .6708 to .7357, and for overall TTR's .1850 to .3932 and .2689 to .4079 for the patients and freshmen, respectively.

4. Group mean segmental TTR's were obtained by averaging the mean segmental TTR's for the individuals within each group. The mean segmental TTR for the schizophrenic group was found to be significantly lower than the mean segmental TTR for the freshmen. The variance of the distribution of mean segmental TTR's for the patients was found to be significantly greater than the variance of the corresponding distribution for the freshmen. When the analysis of the significance of the difference between the group mean segmental TTR's was extended by using *t* to establish limiting values of the true mean for each group, it was found that there was no overlap between these 'confidence intervals' for the two groups at the one per cent level of confidence.

5. Comparisons were made to determine the effect of certain variables, among the schizophrenics, on their mean segmental TTR's. These intra-group comparisons indicated that differences in intelligence test scores, level of educational attainment, and duration of confinement in the hospital had relatively insignificant influence on the TTR's for the patients, and did not adequately account for the differences between the schizophrenic patients as a group and freshman students as a group.

6. Written language samples obtained in this study were compared with spoken language samples obtained by Fairbanks (12) from schizophrenic patients and freshman students. The mean TTR's for both types of subjects run considerably higher for written than for spoken language. This finding may be attributed to the fact that, generally speaking, an individual's written language is a more finished product, permitting more alter-

ing and rearranging of the words used, than is his spoken language.

7. In regard to overall TTR's it was found that the mean overall TTR for the schizophrenic patients was significantly lower than the mean overall TTR for the freshmen, and that the variability in overall TTR's for the patients was significantly greater than the variability in overall TTR's for the freshmen. When t was used to set limiting values of the true mean overall TTR for each group, there was slight overlap in the intervals for the patients and freshmen at the one per cent level of confidence, but there was no overlap in these intervals at the two per cent level of confidence.

8. Differences between the sexes for the two groups in regard to mean segmental TTR's and overall TTR's were not statistically significant, nor was the variability for either sex significantly greater than the variability for the other sex with regard to either of the measures.

9. Correlation between mean segmental and overall TTR's resulted in a Pearson product-moment correlation coefficient of .62 for the patients and .62 for the freshmen. For all subjects the r was .71.

Grammatical Analysis

1. Differences between schizophrenics and freshmen in relative frequency of usage of each of five grammatical classifications (adjectives, adverbs, nouns, pronouns, and verbs), expressed as percentages of the total number of words used, were not statistically significant, with the possible exception of the difference between the groups in relative frequency of usage of nouns, which was significant at the five per cent level of confidence, the patients using more nouns than the freshmen.

2. Differences between males and females within each group in relative frequency of usage of the grammatical categories tested were not statistically significant for schizophrenic patients nor for freshmen.

3. Comparison with Fairbanks' data shows that there is a marked increase in percentage of nouns, adjectives, prepositions, and articles, for both groups, and in conjunctions for schizophrenics, in written over spoken language, and an increase in percentage of pronouns, verbs, adverbs, and interjections, for both groups, and in conjunctions for the freshmen, in spoken over written language.

4. Ratios of adjectives to verbs, adjectives to nouns, and adverbs to verbs were generally higher for the freshmen than for the patients, the difference with regard to the adjective-verb quotient being the greatest; this difference fell very slightly short of significance at the five per cent level.

Type Frequencies

1. Eighty-three words were common to the lists of one hundred most frequently used words for both schizophrenics and freshmen.

2. The number of neologisms, i.e. privately coined words, was fifty-seven for the schizophrenics, and five for the freshmen.

3. When the vocabularies for each group were considered from the point of view of the number of types used to make up a certain per cent of the total number of words, it was found that ten types made up slightly more than twenty-five per cent of the tokens for each group, and that ninety-five types made up fifty per cent of the tokens for the schizophrenics, while ninety-six types made up fifty per cent of the tokens for the freshmen.

4. Differences in the frequencies with which certain types occurred in the

spoken language of schizophrenics and freshmen reported by Fairbanks were not found in the written language of these two groups.

Conclusions

Of the measures used in this study the type-token ratios appear to offer the most fruitful means of differentiating quantitatively written language samples of the type investigated. With the exception of the adjective-verb quotient, and perhaps certain other ratios of parts of speech, the grammatical analysis did not prove useful in this respect. From the results reported by Fairbanks (12) as to the frequency of certain types in spoken language, and from observations of clinical manifestations of ego-centricity, negativism, and frequency of neologisms, the prediction might logically have been made that an investigation of type frequencies would provide a quantitative differentiation of the language of the groups studied. However, the results of the analysis were contrary to this prediction. It is possible that the formality of the writing situation offers a possible explanation of the relative infrequency of self-reference terms, for example, in the written language of schizophrenics. However, the fact still remains that the freshmen students used relatively more first person singular pronouns, while the patients used relatively fewer such pronouns, in written as compared to spoken language. Two other considerations may be mentioned in this respect. It could be postulated that the task assigned the subjects in this study, that of writing a "life story", would tend to increase the frequency of reference to self. This may actually have operated to increase the frequency of self-reference for the freshmen, but for the schizophrenic patients this effect may have been counteracted to a large extent by their tendency to enumerate, and to get 'off the track' in recounting their life histories by describing certain places, events, or things, with little or no reference to their own relation to such places, events, or things. This was particularly noticeable in the writing of some of the patients, one of whom went to great pains to describe how *one* (or *you*) may "bake bread", "can apples", "teach a class in geography", etc., but with almost no reference to *self* involved in such descriptions. It appears obvious from the lack of differentiation between the two groups in terms of the frequency of specific types, that further investigations into this problem will require the formulation of certain other measures designed to offer a means of evaluating the 'adequacy' of the language from a different standpoint.

Insofar as this is a study of 'psychopathological' language on the one hand, and 'normal' language on the other, certain conclusions may be drawn as to the differences between these types of language. The 'normal' subjects investigated in this study appear to have a more highly differentiating language structure in that they use more adjectives per noun, more adverbs per verb, and more adjectives per verb, than do the schizophrenics. This may be interpreted to mean that on the whole they define, modify, and restrict their language in such a way as to make it more accurately representative of the actualities which they are attempting to symbolize. The assumption that 'normal' language structure is more highly differentiated is further substantiated by the fact that the 'normal' subjects have higher type-token ratios indicating that they use more *different* words in producing a given verbal output, than do the schizophrenic patients.

The language of schizophrenics does not appear to be differentiated from 'normal' language in terms of the specific most frequently used words. The vocabu-

laries of the two groups in this study appear to be very similar in that there is an overlap of eighty-three words between the lists of one hundred most frequently used words for each group. The only differentiating feature which a study of the vocabulary pointed to was the relative frequency of neologisms in the language of schizophrenics, as compared to the frequency of their occurrence in the language of freshmen.

As a preliminary investigation this study has provided a quantitative differentiation of language of different types of individuals, and points the way to further research with particular reference to determining the degree of correlation between these measures and other pertinent variables, and to a comprehensive study of language development. Further development and modification of such quantitative measures may provide a means of constructing scaled continua with reference to which any given language sample might be evaluated.

BIBLIOGRAPHY

1. Babcock, H. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, *18*, 107.
2. Balken, E. R., & Masserman, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, *10*, 75-86.
3. Boder, D. P. The adjective-verb-quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, *3*, 310-343.
4. Busemann, A. *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik.* Jena, 1925.
5. Cameron, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, *50*, 1-34.
6. Cameron, N. A study of thinking in senile deterioration and schizophrenic disorganization. *Amer. J. Psychol.*, 1938, *51*, 650-665.
7. Cameron, N. Individual and social factors in the development of graphic symbolization. *J. Psychol.*, 1938, *5*, 165-184.
8. Cameron, N. Functional immaturity in the symbolization of scientifically trained adults. *J. Psychol.*, 1938, *6*, 161-175.
9. Cameron, N. Deterioration and regression in schizophrenic thinking. *J. Abnorm. (Soc.) Psychol.*, 1939, *34*, 265-270.
10. Carroll, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.* 1938, *2*, 370-386.
11. Chotlos, J. W. Studies in Language Behavior: IV. A statistical and comparative analysis of individual written language samples. *Psychol. Monogr.*, 1944, *56*, 75-111.
12. Fairbanks, Helen. Studies in Language Behavior: II. The quantitative differentiation of samples of spoken language. *Psychol. Monogr.*, 1944, *56*, 19-38.
13. Horn, E. *A Basic Writing Vocabulary, 10,000 Words Most Commonly Used in Writing. Univ. Iowa Monogr. in Educ.*, 1924, No. 4, 225 pp.
14. Jersild, A. T., & Ritzman, R. Aspects of language development: the growth of loquacity and vocabulary. *Child Developm.*, 1938, *9*, 243-259.
15. Johnson, W. *Language and Speech Hygiene.* Chicago: Institute of General Semantics, 1939, 10-12.
16. Johnson, W. Studies in Language Behavior: I. A. program of research, *Psychol., Monogr.*, 1944, *56*, 1-15.
17. Mitchell, M. Unpublished research.
18. Piaget, J. *The Language and Thought of the Child.* New York: Harcourt, Brace, 1932, pp. xxiv + 246.
19. Rorschach, H. *Psychodiagnostic, Arbeten zur angewandten Psychiatrie herausgegeben von W. Morgenthaler.* Bern-Leipzig: Bircher.
20. Skinner, B. F. The verbal summator and a method for the study of latent speech. *J. Psychol.*, 1936, *2*, 71-107.
21. Skinner, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, *1*, 71-75.
22. Thorndike, A. L. *The Teacher's Word Book.* New York: Teachers College, Columbia, 1921, 134 + vi pp.
23. *Webster's New International Unabridged Dictionary* (2nd ed.). Springfield, Mass.; G. & C. Merriam Co., 1934.
14. Zipf, G. K. *The Psychobiology of Language.* New York: Houghton Mifflin, 1935, 336 + ix pp.
25. Zipf, G. K. Observations of the possible effect of mental age upon the frequency-distribution of words from the viewpoint of dynamic philology. *J. Psychol.*, 1937, 4, 239-244.

IV. A STATISTICAL AND COMPARATIVE ANALYSIS OF INDIVIDUAL WRITTEN LANGUAGE SAMPLES

IV. A Statistical and Comparative Analysis of Individual Written Language Samples[1]

JOHN W. CHOTLOS
University of Iowa

I. INTRODUCTION

THE PRESENT INVESTIGATION is concerned with the relation of certain language variables to (1) the length of sample from which they are derived and (2) certain psychologically pertinent factors. In general, the language measures employed are based on a count of the number of different words (types) and the relationship of such measures to the total number of words, and to the factors of I.Q., C.A., locality (city, town, rural), and sex. Similar measures based on parts of speech categories and their relationship to I.Q., C.A., locality and sex will be reported. Finally the relationship of the reliability of these measures to the length of samples from which they are derived will be given attention.

Certain previous investigations have been concerned with closely related problems. To begin with, Carroll (3) has presented an equation describing the relation of the number of different words (*D*) to the total number of words (*N*) in a sample of language. A necessary condition to Carroll's formulation of this relationship is that a specified relationship hold between the frequency of a given word in a language sample and its rank in order of decreasing frequencies. Zipf (15) discovered that when he plotted frequency of a word against the number of words having that frequency on logarithmic co-ordinates, the points approximated a straight line except for the few most frequently occurring words. From this fact he formulated the harmonic series law of word distribution, in which he states that the most frequent word in a large sample of language makes up $1/10$ of the sample, the second most frequent word $1/20$ of the sample, the third most frequent word $1/30$ of the sample, etc. This formulation can be put in the form of the equation

$$F = \frac{N}{10R}$$

in which *F* is the frequency of occurrence of any given word in a language sample, *R* is its rank in order of decreasing frequencies, and *N* is the total number of words in the sample.

Skinner (13) has also presented results pertaining to the relationship between *F* and *R*. In analyzing the results obtained from 1,000 responses to his verbal summator, he plotted ranks of words in order of decreasing frequencies (*R*) against frequency (*F*), expressed as a percentage of the total sample, on logarithmic coordinates, and found the points tended to fall on a straight line.

A deviation from linearity was again noted in the more frequently used words. Skinner (13) also reanalyzed the Kent-Rosanoff (8) data on free association response words in the same manner. He found that when the rank order of words in terms of mean frequency per thousand was plotted against mean frequency per thousand on logarithmic coordinates, the resulting curve was approximately linear

[1] This study was done in the Department of Psychology at the State University of Iowa as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy. It is part of a program of research on language behavior. The study was directed by Drs. Wendell Johnson and Don Lewis. Funds and assistance were provided by the Federal Work Projects Administration in connection with Iowa WPA Projects 4892 and 5960.

for the 100 responses most likely to occur. The equation

$$f = \frac{300}{R^{1.29}}$$

where f is the frequency with which a given association will occur in 1,000 responses and R is its rank in terms of mean-frequency per thousand, he finds to be descriptive of the 75 words having the strongest first associations in the Kent-Rosanoff list. He states that this formula is slightly less accurate for the total sample, and his calculated and observed points appear to agree satisfactorily. However, he states that this equation has little practical significance since the frequency and rank of a word must be ascertained before the equation can be used. Nevertheless, he feels that it has an important bearing on theories of language.

Carroll argued that if one accepts the harmonic series law of word frequency distribution, i.e.,

$$F = \frac{N}{KR}$$

where F is the frequency of any word in a language sample, R its rank in order of decreasing frequencies, N the total number of words in the verbal output sample, and K is a constant which is an indirect index of diversity, then it can be demonstrated that the following equation holds:

$$D = \frac{N}{K}(0.423 + K - \log_e N + \log_e K)$$

where D is the number of different words in a sample, N the total number of words in that sample and K is an empirically determined constant. This equation, if it can be shown to be applicable in general, has very important implications with regard to language since, in the first place, if D is known for a specified N, predictions can be made to other N's, and, secondly, the nature of the curve allows a determination of a maximum value of D, a value which can be correlated to a given type of vocabulary of the individual. Carroll tested this equation with a verbal output sample obtained by means of the verbal summator technique, and on several language samples from literature, and found the empirical points to fall very near to the computed curve.

The language samples which formed the protocols of these investigations into the relationship between D and N and between F and R, have been accumulated from different sources and massed into one language sample or are the product of verbally proficient writers. Thorndike (14) has emphasized that if language is to be viewed as behavior, the motivation, backgrounds, the individual characteristics of the writers or speakers must be taken into consideration. He further suggested that the relationship between F and R reported by Zipf may in some measure be a statistical artifact produced by combining language from varied sources, such combination resulting in a loss of individual variation. In the light of this criticism it seems desirable to apply the mathematical formulations of these relationships to samples of language which are the product of but a single individual, in order to test their adequacy more fully.

A second point of interest in the consideration of previous studies is that various attempts have been made to relate the number of different words in a sample to psychologically pertinent factors. Fairbanks (5) working with spoken language and Mann (9) with written language compared superior university freshmen and schizophrenic patients in

terms of the mean percentage of different words per 100-word segment, i.e., the 100-word type-token ratio. Both investigators found the mean type-token ratio for superior freshmen to be significantly greater than for schizophrenic patients, indicating a wider vocabulary range for freshmen than for schizophrenic patients. Fairbanks suggests that for spoken language, there might be a positive correlation between the 100-word type-token ratio and intellectual level. On the other hand, Mann (9) found that differences in intelligence test scores, level of educational attainment, and duration of hospital confinement had relatively little influence on the type-token ratio of patients in terms of accounting for differences between her two groups. In neither study were significant sex differences found. Fossum (6), studying spoken language obtained from junior college students in a regular speech class, found by means of a correlation technique that the 100-word type-token ratio based on 18 segments appeared to be related to parental occupation, correlation of .56, and to speaking rate, correlation of −.45, but not to vocabulary as measured by the Nelson-Denny Reading Test, correlation of .03, nor to intelligence as measured by the percentile score on the Ohio State Psychological Test, correlation of .09. Fossum found no sex differences in type-token ratio measures in his group.

Thirdly, counts of the number of words in parts of speech categories have been made by Fairbanks (5) and Mann (9). They related these counts to other variables under consideration. Fairbanks (5) found that for spoken language there were differences in the use of various parts of speech between her freshmen and schizophrenic groups. The schizophrenics used proportionately more pronouns and verbs and proportionately fewer nouns and articles. On the other hand, for written language, Mann (9) found that the results of the grammatical parts of speech count were not significant, although there seemed to be a tendency for the patients to use more nouns than the freshmen, a result which does not agree with the comparable result obtained by Fairbanks.

Fourthly, Fossum (6) has attempted to relate the reliability of the type-token ratio to the length of spoken language sample. He found the correlation between 100-word type-token ratios for the two halves of 1,800-word samples to be .58 and he estimated by means of the Spearman-Brown prophecy formula that a sample of 14,000 words would be needed to give a reliability coefficient of .95.

Many important facts about language have been reported, and it is of undoubted importance to know whether the relationships already reported will hold for individual language samples. Many investigators have used verbal output samples in which the individual characteristics of the writers and speakers have been lost through massing of the data. In other instances, where individual variations have been of major interest in the investigation, the samples of individuals have been highly selected. It is proposed in this study to investigate language characteristics of individual verbal output samples in which the population sampled will allow for considerable generalization of the results. The question of whether or not individual verbal output samples will bear out the equations descriptive of massed language data is an important one. Furthermore, the relationship between language measures and I.Q., C.A., sex, etc., appears to need further investigation in popula-

tions less highly selected in these characteristics than the ones which have been so far reported.

II. THE PROBLEM

The object of this study may be oriented around the analysis of the number of different words (*D*) as a function of the total number of words (*N*). This relationship may be symbolized by the formula.

$$D = f(N)$$

In this equation it is apparent that we can hold *N* constant and study *D* in relation to other variables, we can study the variation in *D* with concomittant variation in *N*, and finally we can study the variation in *D* and *N* in relation to other variables, such as intelligence test score, age, etc. The basic unit of analysis is the language sample of a single individual. With this introduction the purpose of this study can be summarized in the following statements and questions:

1. To test empirically the equation derived by Carroll, namely,

   $$D = \frac{N}{K}(.425 + K - \log_e N + \log_e K)$$

   where *D* is the number of different words in a sample of length *N* and *K* is an empirical constant to be determined from the data. It is a further purpose to test the assumptions under which this equation was developed.
2. Can the relationship between *D* and *N* be expressed by some empirically determined curve? If such a curve can be determined can the constants in this curve be given any rational meaning?
3. Does *D* for specified *N*'s differentiate I.Q. groups, age groups, location groups, and sex groups? And what is the extent and direction of these differences?
4. Do sections of the language samples, categorized by parts of speech, reveal any relationships or differences which are not apparent in the sample as a whole? How are the parts of speech which go to make up the total sample interrelated?
5. What is the minimum size of sample that can be drawn to reveal the relationships and differences under investigation?

III. SUBJECTS AND PROCEDURES

As part of a remedial education survey, sponsored by the Iowa Child Welfare Research Station and financed by the Federal Work Projects Administration, approximately 1,000 public school children wrote manuscripts of 3,000 words each under conditions to be specified below. The collection and preliminary analysis of these manuscripts was carried out by Work Projects Administration personnel under the supervision of persons with background training in psychology, who had been given special training for this particular assignment.[2] The survey operated for a period of about two years in five counties of the state of Iowa. These counties are distributed throughout the state in such a way that no two counties were adjacent. Each county survey was operated as a unit, coordination being achieved by a state-wide supervisor stationed at the university. In each county all schools, including the one-room rural schools, were invited to participate in the program.

Unit supervisors were instructed to collect 3,000-word language samples for an allotted number of pupils in their respective units. As it took several hours for a child to write the number of words required, extensive cooperation from the school administrators and teachers was

[2] Professor George D. Stoddard served as general director of the survey, of which the language study was a part; Dr. Wendell Johnson was the technical director; Dr. C. Ecco Abermann and Mr. George Wischner served successively as statewide supervisors; and the present writer was the project statistician.

TABLE 1

Factorial design of experimental sample

		Male			Female		
		City	Town	Rural	City	Town	Rural
		***					
I.Q.** 1	Age 1*	2	2	2	2	2	2
	Age 2	2	2	2	2	2	2
	Age 3	2	2	2	2	2	2
I.Q. 2	Age 1	2	2	2	2	2	2
	Age 2	2	2	2	2	2	2
	Age 3	2	2	2	2	2	2
I.Q. 3	Age 1	2	2	2	2	2	2
	Age 2	2	2	2	2	2	2
	Age 3	2	2	2	2	2	2

* Age 1—149 months and under.
Age 2—150 to 179 months.
Age 3—180 months and over.

** I.Q. 1— 89 and under.
I.Q. 2— 90 to 109.
I.Q. 3—110 and over.

*** The numbers refer to the number of randomly selected subjects in the cell.

necessary. The plan called for collecting an equal number of samples from city, town and rural school children, from equal numbers of boys and girls, and an equal number from each grade from four through twelve. Localities with a population of 25,000 or over were called cities, other localities and consolidated schools were considered as town schools, and rural schools of the one-room variety were considered as rural for purposes of this study. Since, as a rule, the one-room rural schools have only eight grades, it was necessary to classify town school pupils who had a rural school background and whose parents were farmers, as rural in order to fill out the rural categories at the older ages. In collecting this sample the pupils were matched by sex for grade, age (within six months), I.Q. (within five I.Q. points) and socio-economic level (within the limits of 1920 U. S. census occupational classification system). No pupil under eight years nor over eighteen years of age was included in the sample.

The writing was done under the supervision of a worker who remained in the classroom throughout the writing session. Writing sessions averaged about forty minutes in length and, on the whole, four or five writing sessions were required for a child to complete his assigned task.

The worker in charge read the following instructions before the children began to write:

> "You are to write about anything you want to write about. Just make it up as you go along. That is, don't write anything you have memorized such as stories or poems. Just start with the first thing you think of and try to keep on writing steadily."

If a child stopped writing for longer than five minutes or complained that he couldn't go on, the worker was instructed not to tell him what to write about, but to tell him to write on whatever he was thinking about. No positive suggestion as to topics was allowed. Legibility of the manuscript was emphasized and speed was not encouraged. Each day's writing was handed in to the monitor at the close of the session. The worker counted the number of words written, entered the count in his record

TABLE 2

Means and standard deviations of distributions of ages in months for the total group and for the main sub-groups

Group	Mean	Standard Deviation
Age Groups		
149 months and under	129.806	12.684
150 to 179 months	164.389	8.473
180 months and over	189.639	7.779
I.Q. Groups		
89 and under	163.917	26.186
90–109	162.083	26.319
110 and over	157.833	26.455
Location Groups		
City	165.194	24.120
Town	159.861	28.136
Rural	158.778	26.469
Sex Groups		
Male	160.796	25.887
Female	161.759	27.020
Total Group	161.278	26.443

TABLE 3

Means and standard deviations of distributions of Otis intelligence test scores (I.Q. units) for the total group and the main sub-groups

Group	Mean	Standard Deviation
Age Groups		
149 months and under	101.278	14.251
150 to 179 months	101.528	14.297
180 months and over	99.028	14.937
I.Q. Groups		
89 and under	83.833	3.670
90–109	101.861	4.461
110 and over	116.139	4.596
Location Groups		
City	101.056	14.012
Town	100.111	14.240
Rural	100.667	13.383
Sex Groups		
Male	100.556	14.453
Female	100.667	13.300
Total Group	100.611	13.888

and dismissed the subject when he had reached the prescribed quota.

From this basic sample of approximately 1,000 language samples, an experimental sample of 108 was selected to conform to the factorial design in Table 1. Since a complete record was available on each child it was possible to sort the larger sample of manuscripts into the fifty-four cells of the design and select at random two subjects for each cell. There was no matching by sex in the experimental sample as was the case in the survey sample. Intelligence was tested by means of the Otis Quick-Scoring Mental Ability Tests. The Alpha test was administered to pupils in grades one through four, the Beta test to pupils in grades five through nine, and the Gamma test to pupils in grades ten through twelve. In classifying the subjects according to the design in Table 1, no distinction was made between the various forms of these three tests. Ages were computed as of the day the children started writing. Criteria for the locality levels of the design are the same as those mentioned above for the collection of the basic sample. The design permits a distribution of 36 subjects at each of three I.Q. levels: (1) 89 and under, (2) 90 to 109, and (3) 110 and over; 36 subjects at each of three age levels: (1) 12 years, 5 months and under, (2) 12 years, 6 months to 14 years, 11 months, and (3) 15 years and over; 36 subjects at each of three locality levels: (1) city, (2) town, and (3) rural; and 54 subjects in each of the sex groups. Furthermore, many combinations of I.Q., age, locality and sex levels are possible.

Means and standard deviations for distributions of I.Q. and age for the total experimental sample and for the main sub-groups, i.e., in terms of I.Q., age, locality and sex, are presented in Tables 2 and 3.

Following the collection of the language samples, the manuscripts were typed and edited. The definition of a word, it should be realized, is crucial in a study of this type. Quite a bit of free-

dom is permitted in defining a unit of language and the way in which the unit is defined is necessarily a condition that is important in connection with any statements made about language phenomena. For this reason the rules for editing the manuscripts are presented in full. These rules define the fundamental language unit better than any formal definition could. The following rules were followed in editing the samples:

1. Type all words exactly as they are written by the subject. Record each correction by writing it in parenthesis after the word for which it is a correction.
2. Correct each misspelling, recording the word as spelled by the subject and writing the correction after it, in accordance with (1) above.
   Classify as a misspelling any word which as spelled by the subject does not constitute a standard English word (current edition of the Century Dictionary to be used as authority) or a recognizable 'slang', nonstandard word (recognizable to the present investigators).
3. Classify as a substitution and correct in accordance with (1) above, any of the following.
   a. Any correctly spelled homonym or an apparently 'intended' word; e.g., "their" substituted for an apparently intended "there", "bare" for "bear", "four" for "for", etc. Judgment in such cases will involve reasonable interpretation of context.
   b. Any correctly spelled non-homonymous substitution which apparently distorts the 'intended' sense; e.g., "of you own" for "of your own", "is would be" for "it would be". Judgment in such cases will, again, involve reasonable interpretation of context.
4. Do not insert any word apparently or obviously omitted by the subject. For example, if the subject writes, "It would fun to play ball," do not insert the word "be" at the point where the subject obviously omitted it.
5. Record slang or non-standard words as written by the subject. When a slang or non-standard word has a standard equivalent, record this equivalent in parenthesis after the slang word; e.g., write "sneaked" in parenthesis as a correction for "snuk". Any slang or nonstandard word having no standard equivalent is to stand as written by the subject and misspellings of such words when recognizable are to be recorded in accordance with (1) above.
6. Any proper name which consists of more than one word is to be counted as one word; e.g., "John Jones" is one word; "East St. Louis" is one word; but "East St. Louis, Illinois" is two words, since they constitute two proper names, the name of a city and the name of a state; "The Chicago and Northwestern Railroad" are three words since (1) "the" is never to be regarded as an integral part of a proper name, always being counted as a separate word and (2) any class-name to which a proper name is attached is to be counted as a separate word; e.g., "railroad", "hotel", "theatre", "street", etc., even in such an example as "the Hotel Roosevelt", "the" and "Hotel" are to be counted as separate words. A proper name is one that designates the sole bearer of the name, as: there is only one "Chicago and Northwestern" railroad, only one "Great Altantic and Pacific" tea company, only one "General Motors" corporation, etc. The names given above in quotes, therefore, are proper names and each is counted as one word. In "A 1940 Multi-Motored Amphibian P45 Boeing Transport", on the other hand, the various words are qualifying adjectives; there are many Transports—Boeing names one type, and it, rather than "Boeing Transport" is a proper name in this case; there are many Boeing Transports, and P45 merely serves as an adjective—there might be P44, P46, etc. Again "Amphibian" serves as an adjective, and so for the terms, "a", "1940" and "multi-motored". In the example of "Dubuque Senior High School", "Dubuque" is an adjective, of course; "Senior High School", however, is not one word in the same sense

that "Chicago and Northwestern" is one word; "Chicago and Northwestern designates the only railroad that goes by that name, but there are thousands of senior high schools; therefore, "senior" and "high" are to be regarded as adjectives; "Dubuque Senior High School" is to be regarded as four words.

"Mrs.", "Mr.", "Miss" and other modes of address are to be counted as separate words and not as integral parts of proper names; e.g., "Mr. John Jones" are two words.

Titles are not integral parts of proper names, but are to be counted as separate words; e.g., "Doctor Jones" are two words; "Senator Hill" or "Professor Smith" are two words.

Abbreviated titles which consist of more than one unit, e.g., M.D., or Ph.D., or unabbreviated titles which consist or more than one word, e.g., "Speaker of the House" or "Dean Emeritus", are to be counted as single words.

7. Any number is to be counted as one word and all figures are to be written or changed to longhand words. "One", "twenty-seven", "one thousand sixteen' are each a single word. Where time is denoted in numbers, it should be counted as a number; e.g., "7:35," write as "seven thirty-five" and count as one word.

   Where street numbers are denoted, write as customarily spoken: e.g., "1220 Harrison Street" write as "twelve twenty Harrison Street".

   When numbers are placed at the beginning of sentences for no obvious reason they are not to be included in the typewritten copy and are not to be counted. For example, in one or two cases it was noted that the sentences had been numbered by the child. Such numbers are not to be counted.

8. Contractions are recorded as written: e.g., "didn't" is not to be changed to "did not"; "didn't" is one word.

9. Record abbreviations as written by the subject, with full term in parenthesis.

10. Hyphenated words properly hyphenated (Century Dictionary to be used as authority) are to be counted as single words; e.g., "hitch-hiker" is one word. Two words improperly hyphenated are to be counted as two words. Corrections in such cases are to be made according to instructions given above.

11. Any two words, as corrected and tabulated, are different unless spelled exactly alike, except:
    a. Plurals and possessives, and contractions involving apostrophes, are to be differentiated even though they are spelled alike.
    b. Any word which begins with a capital letter solely by virtue of its place in the sentence is not to be classified as different from a word spelled as it is in all other respects.

12. All recognizable words are to be counted and tabulated except in the case where some symbol is used to indicate a previously written word. These symbols are not to be counted as that word. For example,

    "John is going to town tonight.
    " " " " school.
    " " " " be late."

    There are nine written words and only nine are to be counted.

13. Sentences are to be left as the child has written them. Do not change the pronoun to agree with the noun in the following example, "Ruth has one side on his paper", nor change the tense in this example, "It is raining last night". Count these as they stand.

After the manuscripts were typed and edited, the types and tokens were recorded and tabulated separately for each manuscript. This tabulation made it possible to abstract the following language measures:

1. *The number of types in any 100-word segment, 500-word segment, 1000-word segment, and in the total 3,000 words.* Thus, 30 measures from 100-word segments, six measures from 500-word segments, three measures from 1,000-word segments and one measure for the total manuscript were computed. For each subject these were averaged to give the mean number of types in 100, 500, and

1,000-word segments, respectively. The mean number of types in a specified segment can be symbolized by *D,* with a subscript to denote the size of the base from which it is computed. The four measures described above can be symbolized by D_{100}, D_{500}, $D_{1,000}$, and $D_{3,000}$.

2. *The type-token ratio for one-hundred-word, five-hundred-word, one-thousand-word and three-thousand-word segments.* The type-token ratio can be defined as the mean percentage of types in any specified segment. If the number of tokens is symbolized by *N,* then the type-token ratio can be symbolized by $\frac{D_i}{N_i}$, where the subscript *i* specifies the size of the sample on which the type-token ratio was based. In our case, the following type-token ratios were computed:

$$R_{100} = \frac{D_{100}}{100} \qquad R_{500} = \frac{D_{500}}{500} \qquad R_{1,000} = \frac{D_{1,000}}{1,000}$$

$$R_{3,000} = \frac{D_{3,000}}{3,000}$$

in which *R* is a symbol for type-token ratio.

R and *D,* as defined above are equivalent measures so long as the number of tokens, *N,* on which they are based is equal for the different individuals in a distribution. However, in instances where a distribution is made up of measures based on a varying *N,* the two measures are not equivalent. In this study, where *R* and *N* gave equivalent measures, *R* was preferred because it made comparison with results of other studies possible.

3. *The cumulative type frequency curve* is obtained by cumulating the types added in each successive 100-word segment. It is the curve that results when D, the number of types, is computed as a function of *N,* the number of tokens. Since each language sample was sectioned into thirty segments, there are thirty points available in the computation of this curve.

4. *The frequency and rank of each type* was computed from the data. The frequency of a type is the number of times it occurred in the language sample. A ranking of the types, in whole number steps, and in order of decreasing frequency, was made. Types of equal frequency were given an average rank number. The numerical position of a word in this sequence is its rank. The frequency of any type is symbolized by F_i and its rank by R_i.

Following the computation of the above language measures, the language sample was split into four subsamples. The division was made on the basis of these parts of speech: nouns, verbs, adjectives, and adverbs. It did not seem profitable to analyze the pronouns, prepositions, conjunctions and articles at this time since these parts of speech are much more limited in the number of available types. An attempt was made to classify the words on a functional rather than on a formal basis. In each instance the function of the word in the context in which it was found determined its classification. For example, the word *run* in "He will *run*" and "She had a *run* in her stocking" are considered as two different types in this classification, the first a verb and the second a noun, whereas in the previously described definition they would be counted as one type according to editing rule number 11. Curme's (2) text, *A Grammar of the English Language,* was used as a reference and final authority in case of doubt. On the basis of this grammatical analysis these additional measures were abstracted.

5. *The number of nounal, verbal, adjectival and adverbal tokens.*
6. *The number of nounal, verbal, adjectival and adverbal types.*
7. *The type-token ratio for nouns, verbs, adjectives and adverbs.* This measure is not equivalent to (2) above since the number of tokens on which the number of types is based is not equal from individual to individual. Usually type-token ratios are not directly comparable unless they are based on the same number of tokens for each individual, but in this instance, it is felt that distributions of type-token ratios

derived from a varying number of tokens can be justifiably used because the total number of tokens is the same in all the manuscripts. If the parts of speech type-token ratios are weighted by their respective number of tokens, their sum will be found to equal *d,* the number of types in the language sample.

8. *Percentage of nounal, verbal, adjectival and adverbal types.* This measure was computed by summing all the nounal, verbal, adjectival and adverbal tokens for each individual and then finding the percentage which the separate nounal, verbal, adjectival and adverbal types are of this total.

IV. RESULTS

RELIABILITY OF DATA

It should be realized that the recording, tabulating, and counting operations in this study were unusually extensive. Attainment of absolute accuracy in a study of this type is both expensive and extremely difficult in a reasonable length of time. Errors were reduced to a minimum by having all operations done twice and by constant supervision of the workers. Further, the procedures for recording, tabulating, and counting the data were so set up as to make possible continuous checking throughout the operations. One set of verbal samples was tabulated independently by two units of the project in order to get some indication of the accuracy of the work. Forty verbal samples comprised this set and the correlation between the two counts of the number of types in each manuscript was .983.

It will be recalled that our data consist of 3,000-word language samples, the individual words of which have been tabulated in such a fashion as to allow for the determination of the number of different words, types, in 100-word segments, or in any segment which is a multiple of 100 words. In order to test the reliability of the type-token ratios the technique of correlating 'split-halves' was employed. The segments of each of the 108 language samples were split into two halves, the first 1,500 words constituting one-half and the last 1,500 words constituting the other. Mean type-token ratios for 100- and 500-word segments were computed for these two halves and the correlation between the halves computed. The product-moment correlation coefficient for the mean type-token ratios for 100-word segments was .829, and for the 500-word segment type-token ratios the correlation coefficient was .826. Since these type-token ratios were computed for only half of the 3,000 words, it is desirable to have some estimate of what the correlation would be if the whole sample were used as a basis for computing the ratios. Such an estimate can be made by means of the Spearman-Brown prophecy formula. Estimated reliability coefficients for the full length of 3,000 words are .906 for the 100-word type-token ratio and .904 for the 500-word ratio. An assumption basic to the use of the Spearman-Brown prophecy formula is that, in this case, the language sample be homogeneous throughout its length in the above two measures. If the two halves are homogeneous, i.e., measure the same aspect of language, then one would expect the mean type-token ratios for the group to be approximately the same for the two halves. These means were found to be 62.58 and 62.64 for the first and last halves, respectively, with regard to the 100-word type-token ratio, and 40.63 and 40.65 for the first and last halves, respectively, with regard to the 500-word type-token ratio. On this basis, the assumption of homogeneity throughout the verbal sample for these two measures appears to be tenable.

It may be, however, that 3,000 words is an insufficient number to reveal any trends in the behavior of these two measures. One long sample of 18,000 words was available and was used to test the hypothesis of homogeneity on a longer language sample. This sample was obtained under the same conditions and rules as the shorter samples, except that the subject volunteered to do the task. Several subjects volunteered to write long samples, but only in this one instance was the task carried beyond the 3,000 word quota. It is recognized that one subject, as such, has no statistical status and that no generalization whatever can be made to the language behavior of other children. It is offered as a particular case of language behavior and because it may be provocative of future leads. The child who wrote this long sample was fifteen years old, attended senior high school, had an I.Q. of 120, and came from a town school. The long sample was sectioned into six 3,000 word sub-samples and the 100, 500, 1,000, and 3,000-word type-token ratios computed for each sub-sample. These results are presented in Table 4. Since the type-token ratios in the first three columns, i.e., for 100, 500 and 1,000 words, are means, it was possible to do an analysis of variance in which one estimate of the population variance is obtained from the deviations of the individual type-token ratios and the other from the variation in the means. None of these three analyses proved significant, the F-ratios being less than one in the case of the 100-word TTR, 1.756 for the 500-word TTR and 1.646 for the 1,000-word TTR. For significance at the five per cent level of confidence, F-ratios of 2.30, 2.54 and 3.20 respectively, would be required. The 3,000-word TTR could not be tested by this technique since there is but one measure of it in each sub-sample.

TABLE 4
Type-token ratio measures of an 18,000-word language sample

3,000-word Sub-sample	Mean 100-word TTR*	Mean 500-word TTR	Mean 1,000-word TTR	Mean 3,000-word TTR
1	71.57**	50.33	42.40	31.20
2	69.13	48.37	40.80	29.57
3	68.87	47.13	40.00	29.27
4	70.13	48.60	41.07	30.37
5	69.83	47.27	38.50	27.40
6	70.40	45.97	36.40	24.10

* TTR is an abbreviation for type-token ratio.
** All type-token ratios are expressed as percentages.

On the basis of this analysis one would infer that the differences between the various sub-sample TTR's for 100, 500, and 1,000 words can be attributed to chance and that the hypothesis of homogeneity has not been discounted. However, this statement holds only if the six sub-samples are randomly selected from a population of such sub-samples. Since there appears to be a downward trend in the magnitude of these type-token ratios with an increase in the number of words written, except in the case of the 100-word type-token ratio, this assumption of randomness may not be fulfilled. If the sub-samples cannot be assumed to be randomly selected, the results of the F-test can be ignored and the data interpreted in

terms of a trend. It is the opinion of this investigator that in this one instance the results presented in Table 4 are indicative of a trend toward a reduction in the use of types with an increase in the number of words written. With the exception of the 100-word TTR, the TTR's may not be considered as homogeneous throughout the length of this sample of language. Factors operating to produce such an effect may be (1) reduction in the number of topics available to the child tending to a greater repetition of types related to topics already discussed, (2) change in motivating conditions, such as loss of interest, boredom, competition with other activities, and (3) an adaptation to the writing situation tending to produce stereotyped behavior. Further, the fact that the 100-word TTR does not show this trend may be indicative of chance factors operating in the other three ratios, or it may indicate that the various type-token ratios are not measuring the same aspect of language.

In any event, it may be considered as demonstrated that for 3,000 words the 100-word and 500-word TTR's are homogeneously distributed throughout the length of the sample, but that, on the basis of one case, this homogeneity may not be assumed to be necessarily present much beyond 3,000 words, except perhaps for the 100-word TTR's.

A problem closely related to that of reliability is posed by this question: "What is the minimum number of words that need to be sampled from one individual to obtain an adequate measure of his language behavior in terms of type-token ratios?" The answer to this question is of more practical interest than theoretical, since the number of words sampled has been, necessarily, arbitrarily determined. If a positive answer can be made to this question much laborious and time-consuming work entailed in language studies may be partially eliminated.

Since no child wrote his full quota of 3,000 words in one day and since there appears to be practically no carry-over in any given child's topics from day to day, it is felt that the first part of a child's output could be considered as relatively independent of his last part. Correlations of TTR's for the first and last part of the sample, based on successively larger numbers of words, should give an indication of the reliability of the TTR as the base number of words is increased. For this purpose the following correlation coefficients were computed for 108 pairs of subjects:

(1) TTR of first 100 words against TTR of last 100 words	$r = .378$
(2) Mean 100-word TTR of first 500 words against mean 100-word TTR of last 500 words.	$r = .669$
(3) Mean 100-word TTR of first 1,500 words against mean 100-word TTR of last 1,500 words.	$r = .826$
(4) First 500-word TTR against last 500-word TTR.	$r = .657$
(5) Mean 500-word TTR for first 1,500 words against mean 500-word TTR of last 1,500 words.	$r = .829$
(6) First 1,000-word TTR against last 1,000-word TTR.	$r = .813$

If it is assumed that the various type-token ratios under consideration are equivalent measures, we see that the 1,000-word TTR is practically as good a measure of the individual's language as the average 100 and 500-word TTR's for 1,500 words. Further, the 1,000-word TTR gives a reliability which compares favorably with that estimated for average 100 and 500-word TTR's based on the full 3,000 words, .813 as compared to .904 and .906. For most purposes, a type-token ration based on 1,000 words will

prove adequate and in some instances TTR's based on only 500 words might prove useful.

Finally we seek to answer the question: "How strongly are the type-token ratios interrelated?" An answer to this question will also give a partial answer to the question: "Do the different type-token ratios employed in this study measure the same aspect of language?" An answer to these questions was sought by intercorrelating the four type-token ratios (all four based on 3,000 words). These correlation coefficients are as follows:

	R_{100}	R_{500}	$R_{1,000}$
R_{500}	.934		
$R_{1,000}$	.870	.948	
$R_{3,000}$	.745	.925	.952

The multiple-correlation coefficient of R_{100}, R_{500}, and $R_{1,000}$ with $R_{3,000}$ is .99. Besides the fact that the four TTR's are highly interrelated it is noted that the greater the difference between the base number of words the less the correlation. Since, for each individual, these TTR's are based on the same language sample, this result is to be anticipated, to some extent, on *a priori* grounds. Further, inspection of the scatter-diagrams of these intercorrelations revealed that the relationships are linear. On these grounds, we would judge the four TTR's to measure essentially the same aspect of language.

GROUP DIFFERENCES IN LANGUAGE MEASURES

Differences between the levels of I.Q., C.A., locality and sex were investigated by means of the analysis of variance technique. The analysis of the factorial design was carried out for the following language measures:

1. Language measures derived from the total sample:
   a. Type-token ratio[3] for 100-word segments.
   b. Type-token ratio for 500-word segments.
   c. Type-token ratio for 1,000-word segments.
   d. Type-token ratio for the total 3,000 words.
2. Language measures derived from sections of the samples categorized by the following parts of speech: nouns, verbs, adjectives, and adverbs. Each of these four sections of the samples were analyzed separately in terms of the following measures:
   a. Number of tokens.
   b. Number of types.
   c. Type-token ratio computed as the total number of types divided by the total number of tokens of each category.
   d. The percentage which the types of each category is of the total number of types in the four categories.

From these measures, a total of twenty analyses of variance was made. In the tables presenting these results the following symbols are used to designate the various levels of the factors under consideration:

1. I.Q. levels
   I_1–I.Q. 89 and under.
   I_2–I.Q. 90 to 109, inclusive.
   I_3–I.Q. 110 and over.
2. Chronological age levels
   A_1–C.A. of 12 years, five months and under
   A_2–C.A. of 12 years, six months to 14 years, 11 months, inclusive.
   A_3—C.A. of 15 years and over.
3. Locality levels
   L_1—City
   L_2–Town
   L_3–Rural
4. Sex
   S_1–Boys
   S_2—Girls

In the twenty applications of the analysis of variance technique, 208 interac-

[3] All type-token ratios are presented as percentages rather than as decimal fractions.

TABLE 5

Summary of results of analysis of variance for 20 language measures

	I.Q.		C.A.		Locality		Sex	
	Level of signif. F-test	Rank order* of means of groups	Level of signif. F-test	Rank order of means of groups	Level of signif. F-test	Rank order of means of groups	Level of signif. F-test	Rank order of means of groups
R_{100}	1%	$I_1I_2I_3$	1%	$A_1A_2A_3$				
R_{500}	1%	$I_1I_2I_3$	1%	$A_1A_2A_3$				
R_{1000}	1%	$I_1I_2I_3$	1%	$A_1A_2A_3$				
R_{3000}	1%	$I_1I_2I_3$	1%	$A_1A_2A_3$				
Number of Tokens								
Nouns	5%	$I_1I_2I_3$			1%	$L_2L_3L_1$		
Verbs					1%	$L_1L_3L_2$	5%	Boys Girls
Adjectives								
Adverbs					5%	$L_1L_3L_3$	5%	Boys Girls
Number of Types								
Nouns	1%	$I_1I_2I_3$						
Verbs	5%	$I_1I_2I_3$						
Adjectives	5%	$I_1I_2I_3$	5%	$A_1A_2A_3$				
Adverbs			1%	$A_1A_3A_2$				
Type-Token Ratio for								
Nouns	1%	$I_1I_2I_3$	5%	$A_1A_2A_3$				
Verbs	5%	$I_1I_2I_3$			5%	$L_3L_3L_1$		
Adjectives	5%	$I_1I_2I_3$	5%	$A_1A_3A_1$				
Adverbs	5%	$I_1I_2I_3$			5%	$L_2L_3L_1$		
Per cent of Total Types								
Nouns								
Verbs	5%	$I_3I_2I_1$						
Adjectives			1%	$A_1A_2A_3$				
Adverbs	1%	$I_3I_2I_1$			5%	$L_1L_2L_3$		

* Rank order of means in terms of increasing magnitude.

Legend: I_1—I.Q. 89 and under; I_2—I.Q. 90 to 109 inclusive; I_3—I.Q. 110 and over.
A_1—C.A. 12-5 years and under; A_2—C.A. 12-5 to 14-11 years inclusive; A_3—C.A. 15 years and over.
L_1—City; L_2—Town; L_3—Rural.

tion variances were computed, and since only three of these were significant, and then only at the five per cent level of confidence, it was felt that, on the grounds that these measures are highly interrelated, it could be safely assumed that there is no interaction among the four factors under consideration as far as these language measures are concerned. By chance one should expect about ten or eleven of the interaction variances to be significant at the five per cent level of confidence.

Because the main purposes of this analysis were exploratory, it was decided that the more conservative error variance estimate should be used to test the main effects. In 19 of the 20 individual analyses, the I x A x L x S variance afforded the more conservative estimate of the error variance and consequently was used as the error term even though the degrees of freedom available for the F-test were much reduced. In the one instance where the I x A x L x S variance was less than the majority of the interaction variances, an error variance was computed by summing all of the sums of squares of the interaction terms and dividing by the sum of the degrees of freedom of all of the interaction terms.

In cases where the F-test was significant, differences between levels were tested by means of Fisher's t-test. The error variance used as a basis for these t-tests was computed as the residual variance after the variation due to the main effects had been deducted from the total variance. This procedure is permissible on the hypothesis of no interaction and since this hypothesis of no interaction seems tenable, the standard error of difference was computed from these residuals because it permitted a greater number of degrees of freedom in evaluating t.

Summary of Results of Analysis of Variance for Each of 20 Language Variables

The evidence garnered from the ap-

plication of the analysis of variance technique to 20 language variables, reveals their capacity to differentiate groups classified according to the factors investigated. A summary in tabular form is presented in Table 5.[4] Considering the fact that the results of the analyses of variance of these 20 variables are rather ponderous, a skeletonized version of these results would appear to be more appropriate than a detailed account. The summary will collect the significant results for each of the pertinent factors.

A. The I.Q. factor. Of the 20 analyses of variance involving the I.Q. factor 14 resulted in significant F-values. Of these, seven are significant at the one per cent level and seven at the five per cent level of confidence. In general, the direction of differences of the means of I.Q. levels for these variables is in a numerical increase in the value of the measure for increases in I.Q. level. The I.Q. factor seems to be the most strongly related to these language measures.

1. Segmental type-token ratios

The 100, 500, 1,000 and 3,000-word type-token ratios all give F-values significant at the one per cent level of confidence. Means of the three I.Q. levels for the segmental type-token ratios are positively related to I.Q. level.

2. Variables dependent on counts of nouns

Three of the four measures derived from counts of nounal types and tokens resulted in significant F-ratios. The type-token ratio and number of types, respectively, are significant at the one per cent level of confidence, while the percentage of nounal types failed to reach either criterion of significance. In each case, differences in means of these measures among I.Q. levels is in favor of an increase in the mean value of the measure with an increase in I.Q. level.

3. Variables dependent upon a count of verbs

Three of the four measures involving counts of verbal types and tokens result in significant F-values, all significant at the five per cent level of confidence. These variables gave significant F-values: number of verbal types, type-token ratio for verbs, and percentage of verbal types. The direction of differences among means of I.Q. levels for the number of types and type-token ratio, respectively, is positively related to I.Q. levels, while the direction of mean differences for the percentage of verbal types is reversed, the low I.Q. group using a greater percentage of verbal types than either of the other two I.Q. groups.

4. Variables dependent upon a count of adjectives

Measures based on a count of the number of adjectival types and tokens result in only two of the four measures giving significant F-ratios for the I.Q. factor. Both of these are significant at the five per cent level of confidence. The direction of mean differences for these two significant measures is positively related to I.Q. level.

5. Variables dependent upon a count of adverbs

Of the four measures involving adverbs, the type-token ratio for adverbs and the percentage of adverbal types gave significant F-ratios, the latter significant at the one per cent level of confidence and the former at the five per cent level. I.Q. levels prove to be positively related to the mean type-token ratio for adverbs and negatively related to the mean percentage of adverbal types.

B. The C.A. factor. Nine of the 20

---

[4] A complete presentation and discussion of the results of the analysis of variance are on file at the S.U.I. library.

language measures, when each is submitted to an analysis of variance in terms of the factorial design, result in a significant F-value for the C.A. factor. Of these nine measures, six give F-values significant at the one per cent level and the remaining three give F-values significant at the five per cent level of confidence. For each of these nine measures, the means of C.A. levels increase in value with an increment in C.A. level; in other words, the older the child the higher the score in terms of these nine measures.

1. Segmental type-token ratios

The four segmental type-token ratios, when submitted to an analysis of variance, result in F-values which, for the C.A. factor, are significant at the one per cent level of confidence in each case. The direction of mean differences is positive, i.e., the older the child the higher the numerical value of the type-token ratios.

2. Variables dependent upon a count of nouns

Only the type-token ratio for nouns resulted in a significant F-value, and this at the five per cent level of confidence. The older children tended to have a greater type-token ratio for nouns.

3. Variables dependent upon a count of verbs

None of the four variables derived from counts of verbal types and tokens results in significant F-values for the C.A. factor.

4. Variables dependent upon a count of adjectives

Three of the four variables involving counts of adjectives prove significant for the C.A. factor. For this factor, the number of adjectival types is significant at the one per cent level. This evidence points to an increase in the use of adjectival tokens and types as the children grow older.

5. Variables dependent upon a count of adverbs

Of the variables in this category only the number of adverbal types gives a significant F-value (at the one per cent level) for the C.A. factor in the analysis of variance. The tendency is for an increase in the use of adverbal types with age, although the two older groups are reversed, but the difference in means between these two groups is not statistically significant.

C. The locality factor. Six of the 20 measures, when submitted to an analysis of variance, gave significant F-ratios for the locality factor. Two of these six are significant at the one per cent level and the remaining four at the five per cent level of confidence. No general trend in differences among the means of the city, town and rural groups was noted.

1. Segmental type-token ratios

Segmental type-token ratios do not differentiate locality groups. None of these four measures gave significant F-ratios for the locality factor.

2. Variables dependent upon a count of nouns

For the locality factor, only one measure, number of nounal tokens, gave a significant F-value (at the one per cent level). The city group uses, on the average, a greater number of nouns than do town or rural groups, while the rural group uses more than do the town group.

3. Variables dependent upon a count of verbs

For the locality factor, the number of verbal tokens and type-token ratio for verbs, respectively, gave significant F-values. Only the number of verbal tokens is significant at the one per cent level of confidence. The rank order of means for number of verbal types is town, city and rural in order of decreasing magnitude, while for the type-token ratio for verbs

the corresponding rank order is city, rural and town.

4. Variables dependent upon a count of adjectives

For the locality factor, measures based on counts of adjectival types and tokens do not produce any significant F-values.

5. Variables dependent upon a count of adverbs

Three of the four variables derived from counts of adverbal types and tokens successfully differentiate locality groups as judged by the F-test. These three measures are the number of adverbal tokens, adverbal type-token ratio, and percentage of adverbal types. All three are significant at the five per cent level of confidence. City groups tend to use the least number of adverbal tokens and types, but have the greatest adverbal type-token ratio. The town group uses the most adverbal tokens while the rural group uses a greater percentage of adverbal types.

D. The sex factor. For the sex factor, only two of the 20 language variables gave significant results in terms of the analysis of variance. These two measures are number of verbal tokens and number of adverbal tokens, respectively, both significant at the five per cent level of confidence. In both instances girls use a greater number of these classes of tokens than do boys.

E. General summary. In general, it may be said that in terms of the language measures employed, the higher the I.Q. and the higher the age level the more highly differentiated is the language structure of the writers. The use of a proportionately greater number of nouns and adjectives characterizes high I.Q. and older age groups, while the use of a proportionately greater number of verbs characterizes the low I.Q. and younger age groups. Adverb usage is not clearly differentiating among these groups.

On the basis of the analysis of variance one would predict that a correlation exists between the type-token ratios and I.Q. score and between type-token ratios and C.A. The correlation of type-token ratios with I.Q. scores might be attenuated by allowing C.A. to be unrestricted and on the other hand the correlation between C.A. and type-token ratios might be attenuated if I.Q. is allowed an unrestricted range due to the possible counteracting influence of these two factors, although there might be reinforcement rather than attenuation. In any event, it is desirable to determine the relationship between I.Q. and type-token ratio with the effect of C.A. reduced to a minimum, and to determine the relationship between C.A. and type-token ratio with the influence of I.Q. minimized. One method of accomplishing this end is to determine the correlation of I.Q. with type-token ratio within C.A. levels and of C.A. with type-token ratio within I.Q. levels. These correlations may be viewed as empirically determined partial correlation coefficients, in the first case with C.A. held constant and in the second with I.Q. held constant. In each case three measures of the partial correlation coefficient are obtained.

On the basis of the assumed equivalence of the four type-token ratios being studied, only the correlations of the 3,000-word type-token ratio with C.A. and I.Q. were computed. The correlation between 3,000-word type-token ratio and I.Q. and between 3,000-word type-token ratio and C.A. within I.Q. and C.A. groups was also computed.[5] These

[5] See Lindquist, E. F. *Statistical Analysis in Educational Research*, Houghton Mifflin Co., 1940, pp. 219-228.

TABLE 6
Table of correlations of 3,000-word type-token ratio with C.A. and I.Q.

Correlation of	*r*	*N*
$R_{3,000}$ With C.A. (I.Q. unrestricted)	.332	108
$R_{3,000}$ With C.A. Within I_1	.463	36
$R_{3,000}$ With C.A. Within I_2	.395	36
$R_{3,000}$ With C.A. Within I_3	.463	36
$R_{3,000}$ With I.Q. (C.A. unrestricted)	.517	108
$R_{3,000}$ With I.Q. Within A_1	.646	36
$R_{3,000}$ With I.Q. Within A_2	.552	36
$R_{3,000}$ With I.Q. Within A_3	.548	36
Within Groups Correlation of		
$R_{3,000}$ With C.A.	.325	108
$R_{3,000}$ With I.Q.	.420	108

correlations are presented in Table 6. In each case the unrestricted correlation along with the correlation within groups as well as the empirically determined partial correlations are presented. A comparison of these two correlations reveals that the unrestricted correlation and the within groups correlations tend to be very much alike in this instance. This result is to be expected since the range in C.A. and I.Q. is reduced relatively much more than is the range of the type-token ratio scores. It would seem that a better index of the strength of the relationship of the 3,000-word type-token ratio to I.Q. and to C.A. is the correlation of 3,000-word type-token ratio with I.Q. within C.A. levels on one hand and 3,000-word type-token ratio with C.A. within I.Q. levels on the other, since in each case counteracting influences are somewhat reduced. However, inasmuch as the locality and sex factors were not significant, the within groups correlations permits the estimation of the significance of the correlation of I.Q. and of C.A. with 3,000-word type-token ratio. With 100 degrees of freedom, both correlations are significant at the one per cent level of confidence when tested by means of Fisher's t-test.[6] The t-value for the within groups correlation of C.A. with 3,000-word type-token ratio is 3.51; the equivalent measure for the within groups correlation of I.Q. with 3,000-word type-token ratio is 4.64.

The positive results of these correlations are indicative of a relationship between type-token ratios and I.Q. and between type-token ratios and C.A., but the relationships are not strong enough to predict a type-token ratio for an individual from knowledge of his age and his I.Q. score. On the assumption that the correlation of C.A. and I.Q. is zero, the multiple correlation coefficient of type-token ratio with C.A. and I.Q. is only .608, a result which suggests that C.A. and I.Q. are not sufficient factors for completely determining the type-token ratio.

CUMULATIVE TYPE FREQUENCY CURVE

In a recent article, Carroll (3) states that the equation

$$D \underset{K}{\overset{N}{=}} -(.423 + K - \log_e N + \log_e) \qquad (1)$$

where *D* is the number of different words in a language sample of length *N*, *N* is the total number of words in that sample, and *K* is an empirically determined constant, held for the language samples he had under investigation. If this formula could be demonstrated to hold generally, it would be a powerful tool in language research.

Carroll deduced equation (1) by means of certain logical and statistical considerations from the following equation,

$$F = \frac{N}{KR^{a}} \qquad (2)$$

where *F* is the frequency with which any given word occurs, *R* is its rank in order

[6] See Lindquist, E. F. *Op. cit.*, pp. 210-211.

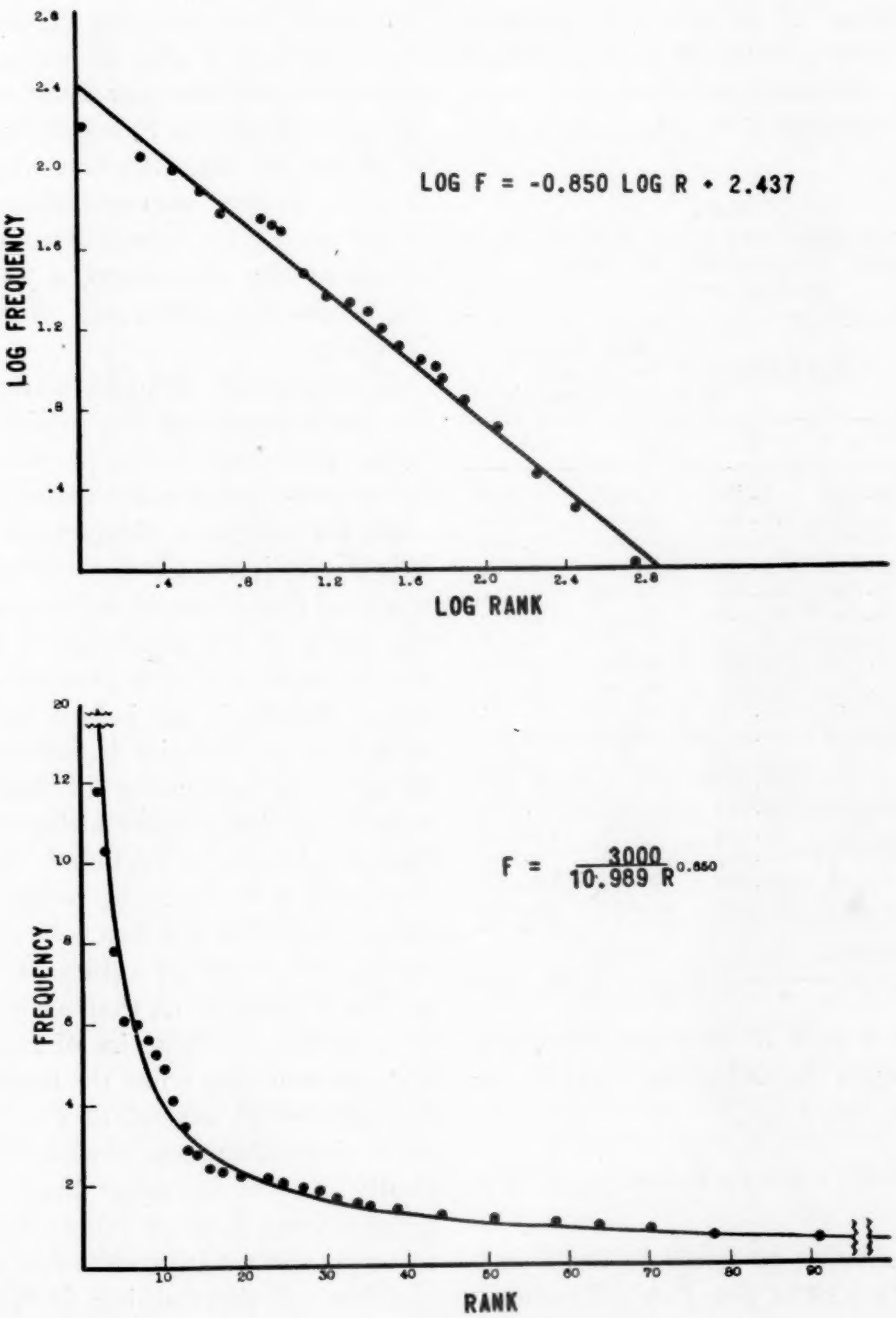


FIG. 1.—Subject no. 12. Graphic representation of the relationship of the rank of a word (R) in decreasing frequency order to the frequency of occurrence (F). The upper plot shows the reduction line in terms of log F and log R. Empirical points are shown in their relation to the curve described by the indicated equation.

of decreasing frequency, N the number of words in the sample from which F and R are computed, and K is an empirically determined constant which has the same meaning as the K in equation (1). A necessary condition to the applicability of equation (1) is that equation (2) hold for the data and particularly, that the exponent of R have a value of 1.0.

To determine whether equation (2) holds for the language samples under

investigation, 18 of the 108 language samples were selected in such a fashion that two randomly selected cases came from each of nine C.A., I.Q. groups. The

TABLE 7
Estimates of parameters, a and K in the fitted equation $F=3{,}000/KR^a$ for eighteen language samples

Subject Number	Total Ranks		First Twenty Ranks Eliminated	
	a	K	a	K
29	0.938	6.716	1.255	2.028
49	0.825	10.549	1.021	4.636
79	0.865	9.080	1.000	5.296
3	0.859	9.709	0.995	5.625
96	0.855	10.211	0.991	5.931
78	0.796	11.700	1.041	4.636
8	0.916	8.513	0.929	8.378
4	0.850	10.989	0.937	7.382
56	0.817	11.646	0.949	6.701
12	0.828	11.619	0.948	6.969
101	0.852	10.142	1.049	4.370
86	0.815	12.948	0.889	9.294
44	0.825	11.891	0.941	7.246
97	0.827	10.741	1.065	3.909
43	0.809	14.495	0.808	14.528
6	0.861	10.003	1.010	5.372
100	0.829	11.299	0.907	8.319
22	0.841	12.448	0.850	12.165
Means	0.845		0.974	

variables *F* and *R* were measured for each sample. Equation (2) can be reduced to

$$\text{Log } F = -a \log R + \log \frac{N}{K}$$

which is seen to be linear in log *F* and log *R*. If a plot of (log *F*, log *R*) can be considered linear, then the line of best fit can be determined by means of the method of least squares. A graphical representation of a representative plot of (log *F*, log *R*) and of (*F*, *R*) along with the best fitting curve is presented in Figure 1. Similar curves were computed for each of the 18 selected samples of language. It can be noted that the fit is not good for the lower ranks, and since Carroll states that equation (2) holds only for ranks greater than about 20, the best fitting straight line was fitted to each of the 18 plots of (log *F*, log *R*) for a series of points in which the first 20 (approximately) points corresponding to the lower 20 ranks were eliminated. Estimates of the parameters, *a* and *K*, for 18 language samples are presented in Table 7.

If equation (1) is to have any generality, then equation (2), which as was noted previously is the harmonic series law of word frequency distribution, must hold for language samples in general. Specifically it must be demonstrated that a plot of (log *F*, log *R*) is linear and that the value of the exponent of *R* is 1.0. On the basis of the 18 plots of values of (log *F*, log *R*), it was judged that the assumption of linearity is reasonable, although the possibility of some other function giving a better fitting reduction line should not be excluded. As for the exponent of *R*, it can be noted that the curves in which the first 20 ranks were eliminated result in values of *a* which are much closer to 1.0 than are the values of *a* for the entire series of ranks. Further, we note that when the first 20 ranks are eliminated, several of the 18 equations give estimates of *a* which are practically 1.0. On the other hand we note quite a range in the *a* values, from 0.808 to 1.235, and we are confronted with the problem of determining if these estimates of *a* are sufficiently close to 1.0 to support the assumption that the value of this parameter is 1.0.

On the assumption that the value of the parameter *a* is 1.0, we can, by means of the t-test test the hypothesis that the mean value of *a* for these 18 language samples differs from 1.0 within the limits of chance. The mean of the 18 *a*-values differs from 1.000 by 0.026 and results

TABLE 8

Table of K-values for eighteen language samples computed at successive five-hundred-word points

Subject Number	500 Words	1,000 Words	1,500 Words	2,000 Words	2,500 Words	3,000 Words	Mean
79	5.69	6.03	6.24	6.45	6.56	6.56	6.24
3	6.49	6.67	6.81	6.86	6.99	7.10	6.82
96	6.94	6.97	7.14	7.22	7.28	7.28	7.14
6	7.14	7.08	7.22	7.29	7.30	7.30	7.22
78	7.13	6.88	6.82	6.88	6.91	6.99	6.94
49	6.50	6.75	6.87	6.87	6.90	6.94	6.81
101	6.66	6.65	6.82	6.96	7.09	7.13	6.89
56	6.90	7.15	7.15	7.28	7.31	7.32	7.19
12	6.59	6.78	7.00	7.17	7.42	7.41	7.06
44	7.47	7.38	7.46	7.39	7.31	7.41	7.40
4	6.92	7.21	7.20	7.37	7.36	7.36	7.24
97	6.88	6.89	6.84	6.87	6.90	6.98	6.89
8	7.08	6.75	6.93	6.99	7.06	7.13	6.99
100	7.43	7.24	7.25	7.36	7.40	7.43	7.35
86	7.47	7.45	7.47	7.57	7.53	7.62	7.52
22	7.33	7.37	7.60	7.81	7.85	7.92	7.65
43	7.91	8.02	7.80	7.93	7.97	8.05	7.95
29	6.17	6.22	6.44	6.48	6.58	6.68	6.43
Means	6.928	6.972	7.059	7.153	7.207	7.256	

in a t-ratio of 1.178, which, with 17 degrees of freedom, gives a probability value greater than 0.2 but less than 0.3, a probability interval which, judged by the ordinary criterion of significance, is not significant. This statistical test is not entirely satisfactory, however, since the assumption we wish to test with reference to the value of the parameter *a* is that it is 1.0 in each individual case and not that the mean of a distribution of randomly selected language samples is 1.0, although the latter is necessarily true if the former holds. On the basis of this test, it is reasonable to accept the hypothesis that the mean value of the parameter *a* might be 1.0, and, in some degree, the assumption of a harmonic series law of word frequency distribution is seen to be validated when approximately the first 20 ranks have been eliminated.

Equation (1) is peculiar in that it contains only one parameter, *K*. An estimate of this parameter can be determined from any one point on the cumulative type-frequency curve. If *K* is constant, as it must be if equation (1) is to hold, then estimates of *K* computed at various points along the cumulative type-frequency curve should differ from one another in a chance fashion. Since we can get a distribution of *K*-values for each individual sample, as well as a distribution of *K*-values at each successive point along the curve for a group of individuals, we can derive two estimates of the population variance, one based on the variance due to the difference in means at successive points and the other a remainder variance computed from the total sums of squares after the variation due to individuals and to successive points along the curve have been deducted. On the assumption that the various estimates of *K* differ only by chance, the F-ratio of the two estimated variances should be non-significant.

For the same 18 language samples used to test equation (2), *K*-values were computed at each successive 500-word point along the cumulative type-frequency curve. Mean *K*-values for the 18 samples

TABLE 9
Results of analysis of variance of K-values for eighteen language samples

Factor	Sums of Squares	d.f.	Variance	F
Individuals	17.0892	17		
Sample size	1.5524	5	0.3105	18.356
Error	1.4377	85	0.0169	
Total	20.0793	107		

Inter-mean K-value differences

	500 Words	1,000 Words	1,500 Words	2,000 Words	2,500 Words
1,000 Words	0.044*				
1,500 Words	0.131	0.087			
2,000 Words	0.225	0.181	0.094		
2,500 Words	0.279	0.235	0.148	0.054	
3,000 Words	0.328	0.284	0.193	0.103	0.049

* Differences greater than 0.0848 are significant at the 5% level of confidence. Differences greater than 0.1116 are significant at the 1% level of confidence.

were also computed at each successive 500-word point as well as the mean *K*-value for each sample. These data are presented in Table 8. It is noted that there appears to be a systematic tendency for the value of *K* to increase as the base number of words from which it is computed increases, although this tendency is not of equal strength in all cases.[7]

The data in Table 8 were subjected to an analysis of variance in order to determine if the variation in mean *K*-values at successive 500-word points could, statistically, be allocated to chance factors. The error variance used to test the sample size variance, i.e., variation derived from the means at successive 500-word points, was the interaction variance of individuals and sample size. The results of this analysis of variance are presented in Table 9. The F-ratio of 18.356 is significant at the one per cent level of confidence. This result may be interpreted as meaning that the mean *K*-values computed at successive 500-word points along the curve cannot be considered as representing populations which are equally variable or which have equal means. On the basis of a significant F-test, inter-mean *K*-values were tested by means of the t-test. Of the 15 differences among the six means, 12 give a t-value significant at least at the five per cent level and nine give t-values significant at the one per cent level. If the means of the *K*-values differed by chance only, one would expect less than one of the 15 differences to be significant at the five per cent level when tested by means of the t-test.

A further test of the adequacy of the hypothesis that *K* is constant is afforded by an analysis of the behavior of *K* in the 18,000-word sample. *K*-values were computed at successive 1,000-word points throughout the 18,000 words. Since *K* cannot be validly estimated at very large values of *N* it was necessary to determine whether or not *K* could be validly computed at each of the successive 1,000-word points along this curve. In equation (1), *D* is a double-valued function of *N*, i.e., there are two values of *N* which will satisfy the equation for any given *D*. Disregarding the portion of the curve described by equation (1) for nega-

[7] In one case, subject number 79, the value of D at $N = 3{,}000$ was not large enough for a valid computation of *K*. In order to complete the design the value of *K* computed at 2,500 words was used as the best estimate of *K* at $N = 3{,}000$.

tive values of N, the curve may be said to have its origin at point (0.0), to rise to a maximum and to fall indefinitely for all values of N beyond this maximum. The usable portion of the curve is from its origin to its maximum. The maximum point on the curve is determined from K and for each value of K a maximum N can be computed beyond which the value of D computed from equation (1) is not valid. If for each K there is a maximum N, then for each value of N there is a minimum value of K which is valid for that value of N. These minimum values of K for specified N's can be determined by setting the first derivative of equation (1) equal to zero and solving for K. The first derivative of equation (1) is

$$\frac{dD}{dN} = \frac{1}{K}(K + \log_e K - \log_e N - 0.577) \quad (4)$$

and setting

$$\frac{dD}{dN} = 0$$

$$\operatorname{Log}_e N = K + \log_e K - 0.577 \quad (5)$$

as the equation from which we can determine the maximum point on the curve for specified values of N. In terms of the independent variable, N, the limits of the usable portion of the curve derived from equation (1) are from $N = 0$ to

$$N = E_{(K + \log_e K - 0.577)}. \quad (6)$$

From equation (5) minimum values of K, which for a specified N give a maximum, can be solved. From this value of K, then, the minimum D value for the specified N and K can be computed from equation (1). These minimum D and K values, along with the empirically determined K-values at each successive 1,000-word point of the cumulative type-frequency curve, are presented in Table 10.

Table 10

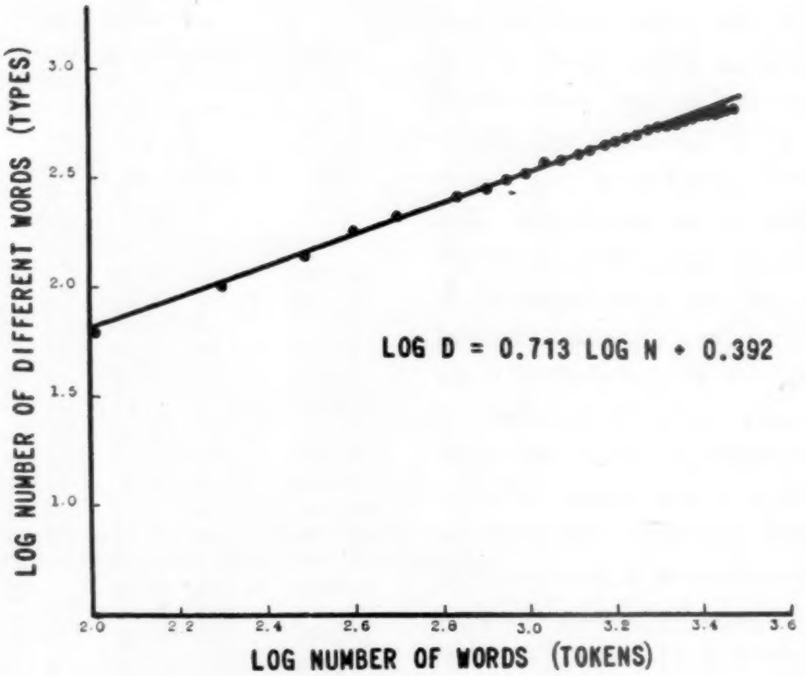
K-values computed at successive 1,000-word points of the 18,000 word sample

N	D	K	Minimum K for given N	Minimum D for which K is valid
1,000	434	7.82	5.73	175
2,000	732	8.03	6.35	316
3,000	936	8.00	6.68	449
4,000	1,175	8.18	6.94	576
5,000	1,376	8.27	7.13	701
6,000	1,532	8.27	7.29	823
7,000	1,700	8.34	7.42	943
8,000	1,889	8.42	7.54	1,061
9,000	2,021	8.45	7.64	1,178
10,000	2,193	8.51	7.74	1,292
11,000	2,344	8.56	7.83	1,405
12,000	2,465	8.58	7.90	1,519
13,000	2,585	8.62	7.97	1,631
14,000	2,703	8.64	8.04	1,741
15,000	2,786	8.64	8.10	1,852
16,000	2,884	8.66	8.16	1,961
17,000	2,966	8.66	8.21	2,071
18,000	3,025	8.67	8.26	2,179

Apparently K-values can be validly computed throughout the length of this sample. However, it is to be noted that if we were making a prediction of D from 3,000 words, the number of words collected from the other subjects, the estimate of the maximum number of different words in this child's vocabulary would fall somewhere between 2,585 and 2,703 words, i.e., at the point along the N-axis beyond which computation of D from $K = 8.00$ is no longer valid. The data do not justify the prediction that the child's vocabulary is exhausted after writing between 13,000 and 14,000 words, as new words are being added throughout the length of the sample.

On the other hand, it is noted that beyond N equal to about 11,000 or 12,000 the value of K remains fairly stable. This sample, like the previously discussed samples, gives estimates of K which show a systematic tendency to increase for successive values of N.

The evidence points to a rejection of the hypothesis that, for these language samples, K is a constant. Indirectly, it



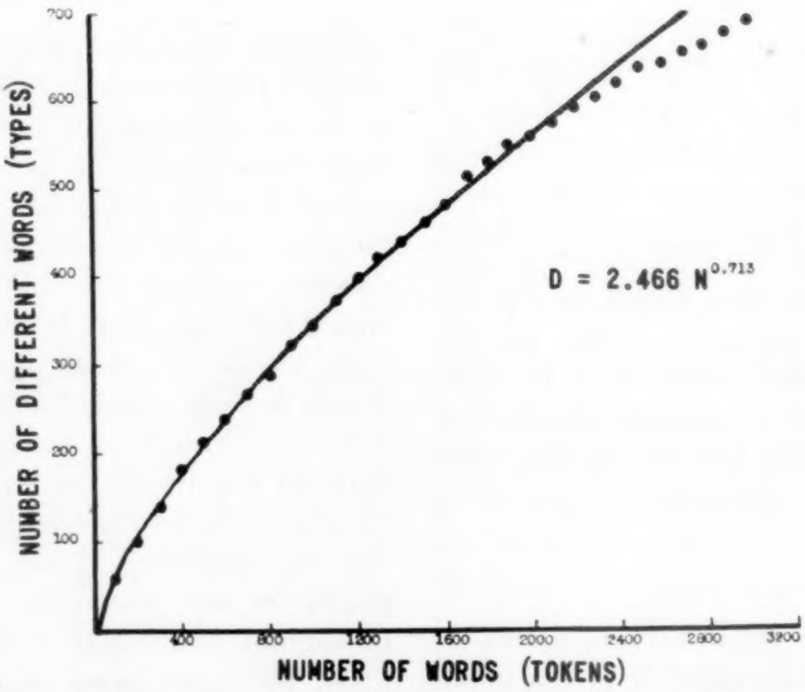


FIG. 2.—A graphic representation of the relationship of the number of different words (D) to the sample size (N). The upper curve is the reduction plot in terms of log D and log N. Plotted from the data computed from language sample written by subject no. 96. Empirical points are shown in their relation to the curve described by the indicated equation.

means that equation (1) will not adequately describe the data derived from these samples and that D will vary systematically from the obtained values of D. We may, on this evidence, reject the generality of equation (1) for language samples like those used in this investigation, with the qualification that 3,000 words may be an insufficient number of words on which to base an estimate of K for prediction of D's.

An attempt was made to find an em-

TABLE 11

Estimates of parameters, a and b, in the fitted function $D=bN^a$ for 18 language samples

Subject Number	a	b
29	0.618	3.373
49	0.657	3.192
79	0.702	1.656
3	0.725	1.963
96	0.713	2.466
78	0.600	5.023
8	0.646	3.733
4	0.740	2.109
56	0.710	2.570
12	0.743	1.945
101	0.683	2.754
86	0.722	2.606
44	0.690	3.170
97	0.599	4.989
43	0.729	2.786
6	0.679	3.221
100	0.665	3.750
22	0.794	1.614

TABLE 12

Results of analysis of variance of parameter a for groups of I.Q., C.A. and type-token ratio

Factor	Sums of Squares	d.f.	Variance	F
I.Q.	0.004449	2	0.002225	1.642
C.A.	0.010095	2	0.005048	3.723
TTR	0.015547	1	0.015547	11.474*
Error	0.016261	12	0.001355	
Total	0.046352	17		

Factor	Mean	Difference	t
I.Q.			
89 and under	0.669		
90 to 109 (inc.)	0.707		
110 and over	0.693		
C.A.			
149 mo. and under	0.656		
150 to 179 mo. (inc.)	0.709		
180 mo. and over	0.704		
TTR for 3,000 words			
Less than .231	0.660		
Greater than .231	0.719	0.059	3.411*

* Significant at the one per cent level of confidence.

pirically fitted equation to represent the cumulative type-frequency curves. Since it is reasonable to assume that, for any individual, there is a limit to the number of types at his command, it was felt that an equation of the hyperbolic form would best agree with the character of the phenomena at hand, inasmuch as the hyperbolic curves are characterized by asymptotes, which can be correlated to the limit of the writing vocabulary of the individuals in similar situations. However, it was found that the data could not be satisfactorily reduced to a linear form of the hyperbolic curve, at least not a simple hyperbolic curve with two parameters. Possible reasons for this failure will be discussed later.

Of the attempts to fit curves with simple equations to these data, only a plot of log D and log N resulted in what could be considered a linear relationship. The resulting linear function is of the form

$$\log D = a \log N + \log b$$

in which a and b are empirically determined constants. If the above equation holds, then log D is a linear function of log N, and D is a power function of N, of the general form

$$D = bN^a.$$

Equations of this form were fitted to the cumulative type-frequency data for the 18 language samples used in fitting equation (2). Estimates of parameters a and b arrived at by means of a least squares solution of the (log N, log D) plot, are presented in Table 11. A graphical representation of this relationship for a typical plot of (log N, log D) and (N, D) is shown in Figure 2. The curve presented in Figure 2, is typical of the other fitted curves in that the fit for larger values of N is not satisfactory and would make prediction beyond the limits of these data rather hazardous.

In order to determine what relationship exists between the estimates of these parameters and the factors of I.Q., C.A. and 3,000-word segmental type-token ra-

TABLE 13
Results of analysis of variance of estimates of parameter b for groups of I.Q., C.A., and type-token ratio

Factor	Sums of Squares	d.f.	Variance	F
I.Q.	1.211853	2	0.605927	—
C.A.	1.220563	2	0.610284	—
TTR	1.606827	1	1.606827	—
Error	12.739341	12	1.061612	
	—	—		
	16.778584	17		

Factor	Mean
I.Q.	
89 and under	2.946
90 to 109, inclusive	2.620
110 and over	3.255
C.A.	
149 months and under	3.279
150 to 179 months, inclusive	2.646
180 months and over	2.895
TTR for 3,000 words	
0.231 and less	3.239
0.232 and over	2.641

TABLE 14
Estimate of parameters, a and b in the equation $D = bN^a$ for 3,000-word sections and for the total sample of the 18,000-word language sample

Sample Section	a	b
Total 18,000 Words	0.697	3.499
First 3,000 Words	0.729	2.825
Second 3,000 Words	0.752	2.239
Third 3,000 Words	0.785	1.694
Fourth 3,000 Words	0.752	2.193
Fifth 3,000 Words	0.712	2.884
Sixth 3,000 Words	0.680	3.236

tio, three levels each of I.Q. and C.A. as previously defined and two groups, a low and a high, categorized according to the magnitude of the 3,000-word segmental type-token ratio, were subjected to an analysis of variance. The results of the analysis of variance of these two parameters for these groups is presented in Tables 12 and 13. The results of the analysis for parameter *a* indicates that only the type-token ratio factor results in a significant F-value. This result is to be expected since the type-token ratio is a function of *D,* one of the variables in the equation. Inasmuch as the constant *a* determines the rate at which new words are added, it is not surprising that the difference in means between the low and high 3,000-word TTR groups should result in a significant difference, when tested by means of the t-test, in favor of a greater magnitude in the value of *a* for the group with a larger type-token ratio. The I.Q. and C.A. factors show no systematic differences between means of the various groups that are of great enough magnitude to give significant F-values. Since these factors have been shown to be associated in systematic manner to the 3,000-word segmental type-token ratio the marked differences between the two lower levels of I.Q. and C.A. might presumably be attributed to this association.

The results of the analysis of variance of estimates of parameter *b* for these groups indicate no significant factors. Apparently the exponent of *N,* i.e., the parameter *a,* is more influential in determining the differentiating characteristic of the curves than is the co-efficient of *N,* i.e., the constant *b.*

In order to determine if the power function will hold beyond 3,000 words, a curve was fitted to the data of the 18,000-word sample. The sample was divided into six 3,000-word samples and a curve fitted to each 3,000-word section as well as to the total sample. Estimates of the parameters of the power function for the total sample and for each section is presented in Table 14 and a graphical representation of the relationship for the 18,000 words is presented in Figure 3. Again, it is noted that the fit for the larger values of *N,* beyond *N* equal about 14,000, is poor. The empirical points diverge considerably from the curve.

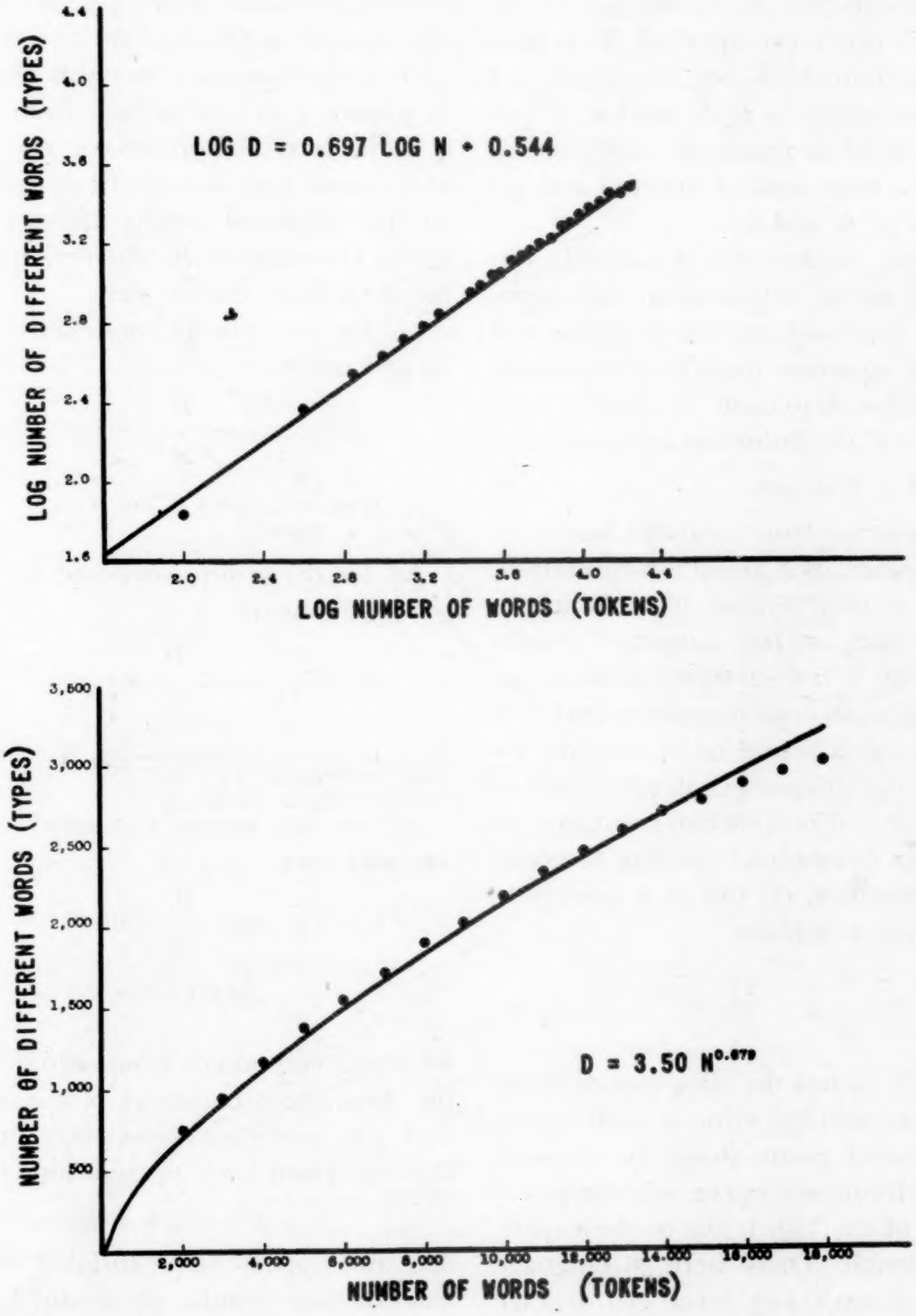


FIG. 3.—A graphic representation of the relationship of the number of different words (D) to the sample size (N) and of log D to log N for the 18,000-word language sample. Empirical points are shown in their relation to the curve described by the indicated equation.

A second empirically fitted curve was derived by generalizing equation (1). A transformation of equation (1) can be made by writing the function as follows,

$$\log_e N = -K\left(\frac{D}{N}\right) + (\log_e K + K + 0.423), \quad (7)$$

which, since the terms in the right-hand bracket are all constants can be written as,

$$\log_e N = -K\left(\frac{D}{N}\right) + C. \quad (8)$$

Equation (8) is a more general function,

of which equation (7) is one special instance. In this form equation (8) is seen to be linear in $(D/N, \log_e N)$. Thus, if a plot of the variables D/N and $\log_e N$ can be considered as linear, we can proceed to make a least squares solution and get estimates of K and C.

We note, further that if Carroll's formulation of the relationship, i.e., equation (1), is to hold the two parameters of the fitted equation must have a definite relationship. Equation (1) makes the parameter C the following function of K,

$$C = \log_e K + K + 0.423. \tag{9}$$

In this way we have a further test as to the adequacy of Carroll's formulation, one that is more satisfactory since it uses all the data in the language sample rather than a few selected points along the cumulative type frequency curve. If equation (9) is found to be tenable, we have direct evidence in substantiation of equation (1); if not, we can substitute an expirically determined curve of the same type as equation (1) but of a more general nature, as follows:

$$D = \frac{N}{K}(C - \log_e N). \tag{10}$$

In order to test the adequacy of equation (8), an average value at each successive 100-word point along the cumulative type-frequency curve was computed for each of the I.Q. levels of the experimental design. There were 36 language samples at each I.Q. level. An average series of points was felt desirable in order to give the empirical curve a greater stability at each point and also to smooth out chance fluctuations along the curve. The I.Q. levels were chosen because the variable D more clearly differentiated I.Q. levels than C.A., locality or sex levels.

A graphical representation of the relationship between (D/N, $\log_e$ N) and (D, N), as well as the best fitting curve for each of the three I.Q. levels, is presented in Figure 4. It can be seen from Figure 4 that the reduction curves are reasonably linear, and that the fit of the curves to the empirical points appears to be good. The empirically derived equations for these three curves are:

(1) for the group composed of I.Q. 89 and under,

$$\log_e N = -9.910\frac{D}{N} + 9.853 \quad \text{or}$$

$$D = \frac{N}{9.910}(9.853 - \log_e N)$$

(2) for the group composed of I.Q. 90 to 109, inclusive,

$$\log_e N = -10.081\frac{D}{N} + 10.272 \quad \text{or}$$

$$D = \frac{N}{10.081}(10.272 - \log_e N)$$

(3) for the group composed of I.Q. 110 and over,

$$\log_e N = -9.551\frac{D}{N} + 10.321 \quad \text{or}$$

$$D = \frac{N}{9.551}(10.321 - \log_e N)$$

K-values computed from equation (1) for these three curves at $N = 3{,}000$ are 6.97, 7.24, and 7.40, respectively. Further, the restriction basic to equation (1), that

$$C = K + \log_e K + 0.423$$

does not appear to be fulfilled. On the basis of these results, we would be compelled to reject the generality of equation (1) to these data.

A plot of (D/N, $\log_e$ N) was also made for the 18,000-word sample. The plot of (D/N, $\log_e$ N) and D, N) as well as the best fitting curves in each instance, is presented graphically in Figure 5. Apparently the function described above represents the data reasonably well

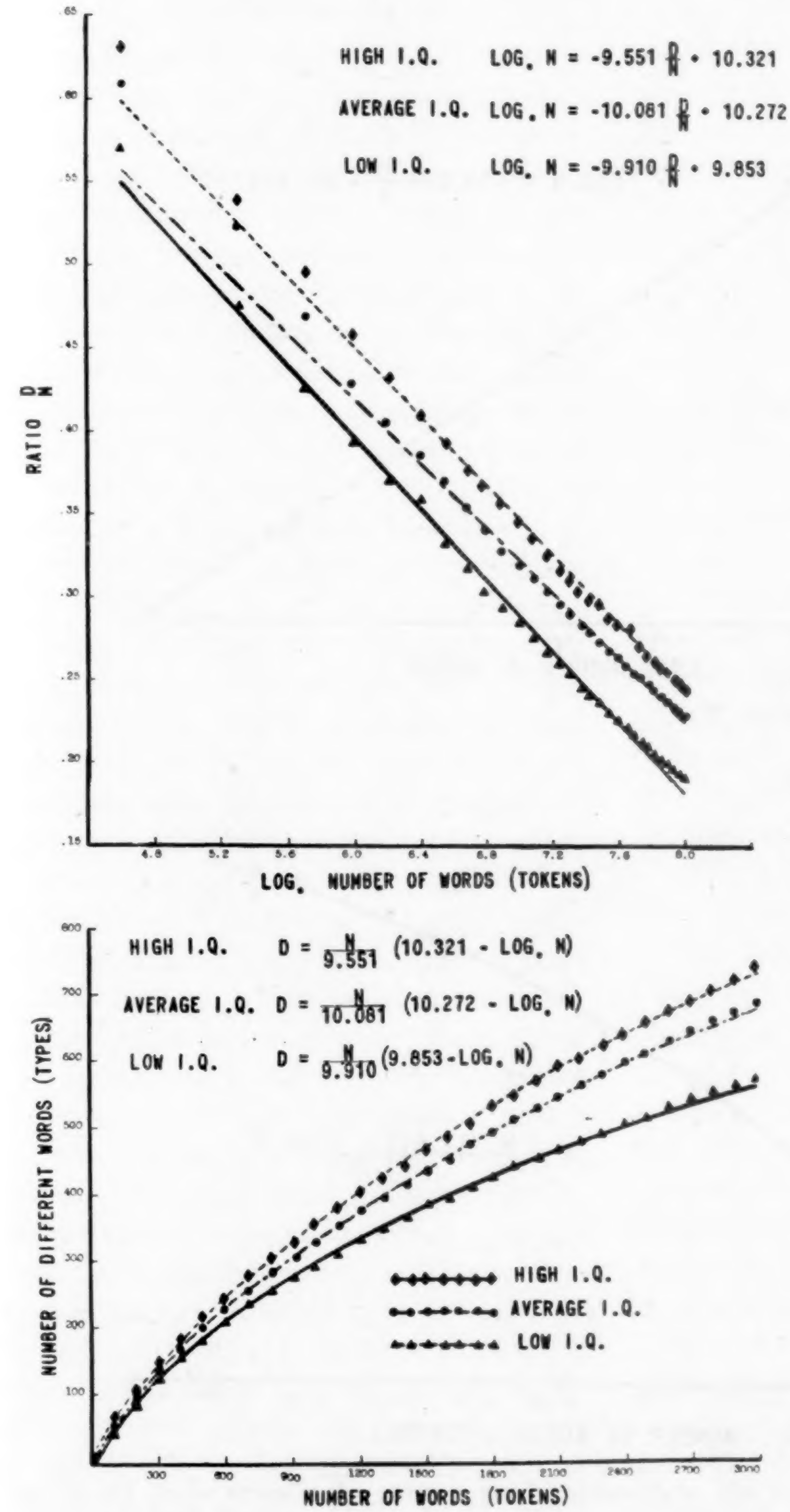


FIG. 4.—Graphic representation of the relationship of the number of different words (D) to the language sample size (N) and of the ratio D/N to $\log_e N$. Plotted for each I.Q. level by averaging D at each successive 100-word point along the N-axis for each of the 36 subjects in that I.Q. level. Empirical points are shown in their relation to the curve described by the indicated equation.

throughout the 18,000 words. The empirically determined equation for this long sample is,

$$\log_e N = -11.268 \frac{D}{N} + 11.670 \quad \text{or}$$

$$D = \frac{N}{11.268}(11.670 - \log_e N)$$

Equations (1) and (10) have similar properties. Inasmuch as in equation (1) the equivalent of parameter *C* is a function of *K*, the maximum point on the curve is also a function of *K*. However, in the empirically determined curve, the parameter *C* is independent of the con-

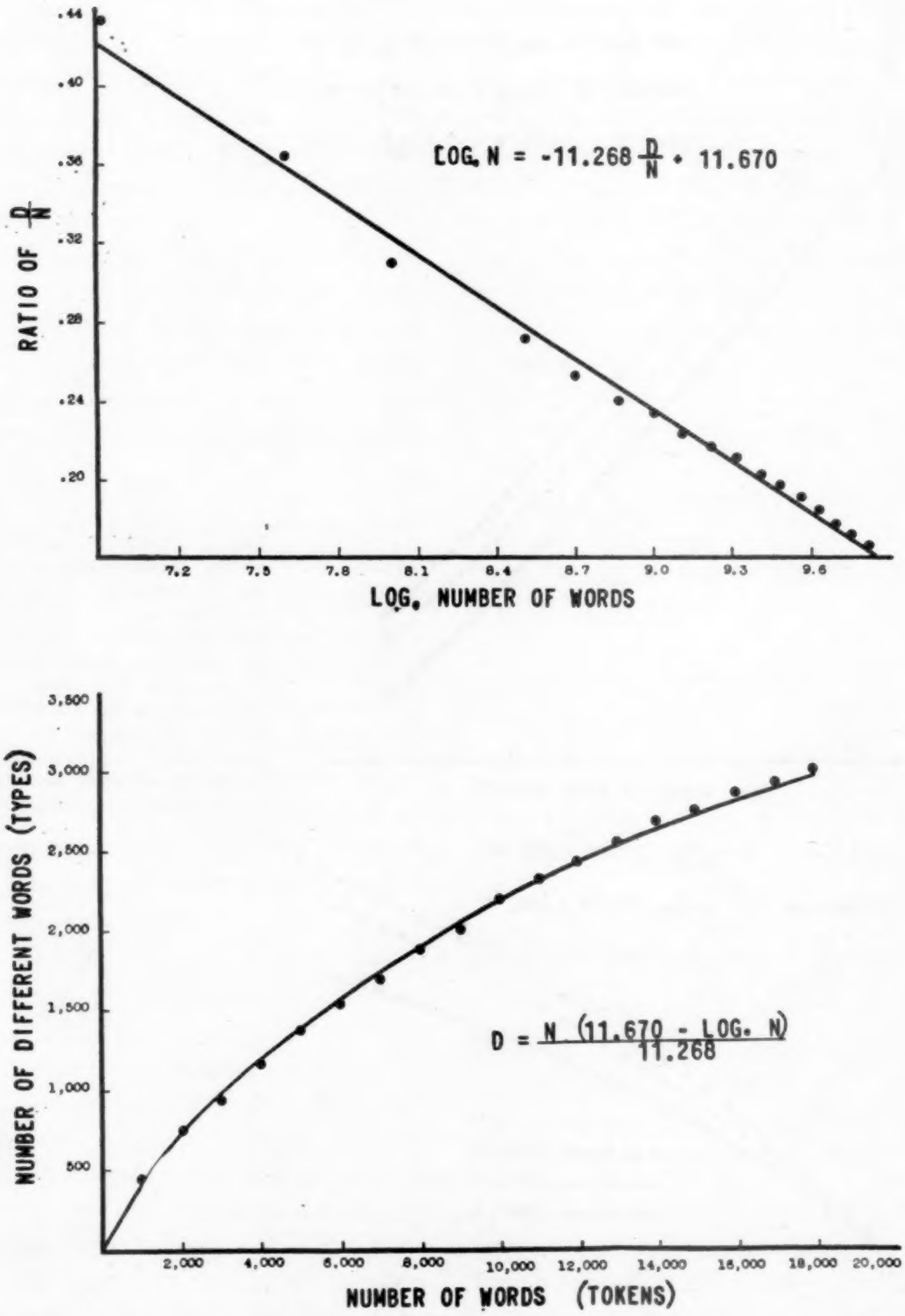


FIG. 5.—A graphic representation of the relationship of the number of different words (D) to the language sample size (N) and of the ratio D/N to $\log_e N$ from the data computed from the 18,000-word language sample. Empirical points are shown in their relation to the curve described by the indicated equation.

stant K, and it is noted that the maximum point on the curve is a function of C and not K. For any given curve of this form, the maximum point on the curve is reached when

$$N = E_{(C-1)} \quad \text{or}$$
$$\log_e N = C - 1$$

If the value of N is greater than $E_{(C-1)}$, computations of D for these values of N are spurious.

V. DISCUSSION

Crucial to the interpretation of the results of this study is the manner of selection and definition of a unit of language. Any one of several language units might have been chosen—syllable, word, phrase, or clause, among others. The simplest and least ambiguous unit to define is the word. For this among other reasons, the unit of language used in this study is the word, and it is generally referred to as a *token* in order to differentiate it from a unit of vocabulary, the *type*. This definition of a language unit is in conformance with that used by Zipf (15), Carroll (3), Fairbanks (5), Mann (9), and Fossum (6) in studies which are comparable to this one. The word, as a unit of language, results in a statistical measure which involves fundamentally simple enumeration. All words are given equal weight in the determination of relationships. It may be questioned whether such a definition of language is entirely satisfactory, since it is known that a large proportion of words in connected discourse must necessarily be made up of the structural or interstitial words which represent relationships among other words and, in and of themselves, carry no meaning independent of the immediate verbal context in which they appear. On this basis, any sample of language may be divided into at least two classes of words, (1) the structural words and (2) the content words. It is possible, though less practicable, to give each word a weight in accordance with its classification in the above terms, or in accordance with its frequency of use in some standard language sample, and thus derive a more suitable measure of language. In any event, the procedure used will, to some degree, affect the character of the results obtained.

The classification of word units according to certain rules and the counting of the resulting classes results in a measure of the number of types, or what may be appropriately called the vocabulary of the sample. Here again we note that our measure is a statistical one, in that it involves classification and enumeration. Once more, each type is given an equal weight in arriving at the numerical value of the measure, regardless of the function each type plays in the language structure. In this instance, however, greater liberty is given us in setting up our classes. What is psychologically the more fruitful method of selecting these classes can only be speculated upon. It might be argued that *'fall', 'fall in', 'fall out', 'fall short', 'fall apart'* should each be classified as a unit rather than as two units on the grounds that each represents a unitary symbol. Such an argument is undoubtedly valid, and if such a procedure were followed it might conceivably alter the results.

The language measures used in this study represent but a fraction of those already in use or that have been suggested for use. Sanford (11) in a recently published investigation demonstrated the utility of some 234 language measures which he used to describe personality differences between two individuals whose language he investigated. Busemann (2) has suggested and used the adjective-verb quotient for the study of personality. Boder (1), following Busemann's suggestion, has also employed the adjective-verb quotient, this time in the study of various types of literature. Mann (9) also applied this measure in comparing schizophrenic patients with university freshmen; she also employed an adjective-noun quotient and an adverb-verb quotient. Johnson (7) has suggested several language measures, among which are included the ones used in this

study. Measures of language, not in terms of word counts, but in terms of themes and topics have been investigated by Skinner and his co-workers (12). This enumeration will give the reader at least a suggestion of the abundance of language measures already in use or proposed. Many others, of course, are possible. The ones used in the present study by no means exhaust the possibilities of language analysis.

Further qualifications of the present results arise with regard to the fact that pertinent determining factors may have been omitted from consideration. Two such factors may be (1) number of topics discussed by each individual and (2) rate of verbal output per unit of time. With regard to the first of the above factors, if we are permitted to assume that, other conditions being equal, the number of types in a sample of language is positively correlated to the number of topics discussed in that sample, then it may appear plausible that, insofar as the more intelligent individuals can sustain a discussion on one topic for a greater number of words and, thus, for a specified number of words discuss fewer topics than less intelligent individuals, who of a necessity must shift topics more frequently to write a given quota, the relationship of intelligence to the number of types will be somewhat attenuated unless the number of topics is given some weight in the determination of this relationship. Again, it seems likely that age differences may be accentuated due to a wider range of interests, ambitions, opportunities, etc. of the older children. The behavior of these language measures within selected topics or fields of writing, as for instance in fiction and scientific writing, may be profitably investigated. In any event, consideration of this factor of number and type of topic, perhaps by an analysis of co-variance technique, will broaden our understanding of language in terms of these measures.

The second factor mentioned above, namely, rate of verbal output per unit of time, might be investigated by somehow weighting the number of types produced by reference to the rate of verbal output. The analysis of co-variance technique is an appropriate method of carrying out such a weighting. Fossum (6), for spoken language, reported a negative correlation of $-.45$ of 100-word type-token ratio with rate of verbal output per unit of time. Whether the relationship, if any, is in the same direction for written language is yet to be determined.

These two above-mentioned factors do not exhaust all of the possibly pertinent factors which may need to be controlled, although they do offer perhaps the greatest promise of successful manipulation. Psychological factors such as motivation, interests, attitudes, ambitions, emotional states, etc. which are admittedly more difficult to control, nevertheless may prove to be significant determining factors in differentiating individuals in terms of these language measures. It was noted that in some instances children who expressed a dislike for the task of writing 3,000 words seemed to show a tendency to write in short jerky sentences with many of the sentences beginning with the same pattern of words, producing a stereotyped effect. On the other hand, children who expressed interest and a liking for this task tended to keep their discussion varied throughout the manuscript and thus probably produced a greater number of types than they would have if the motivation had been less adequate. For example, the boy who produced the greatest number of types in the present group of chil-

dren expressed a desire to be a writer of western fiction. Additional factors that may need to be given consideration in studying language in terms of these measures are: (1) differential effects of fatigue; (2) physiological conditions; (3) socio-economic background in the immediate family sense; (4) season of the year; (5) personality variables; etc. This enumeration certainly does not include all the possibly influential factors and serves only to demonstrate the complexity of the problem.

One implication of the results of this study is that the segmental type-token ratios based on successively larger segments are for practical purposes equivalent measures, insofar as they differentiate among individuals. This implication is inferred from the fact that the intercorrelation of these segmental type-token ratios are uniformly high and linear. However, there are instances in which the value of the type-token ratio for 100-word segments places individuals near the top of the group for this measure while for these same individuals the value of the 3,000-word type-token ratio places them near the bottom of the group. Such individuals would seem to be making efficient use of the vocabulary available to them, and if such an interpretation is justified, development of the notion of efficiency of vocabulary usage should result in fruitful research.

Certain implications arise from the curve-fitting aspects of the present study. In general, tests of the applicability of the equation presented by Carroll (3) point toward a rejection of the generality of this equation to these data. One possible reason for this may lie in the fact that the language samples used in this investigation were much different in certain characteristics from the language samples used by Carroll. The main points of difference are: (1) one type of language sample used by Carroll consisted of the verbal output of a group of subjects which was combined into one unit, while in this study each language sample represents the performance of but one child. It may well be that the aforementioned equation will hold for the first type of language sample but not for the second; (2) the other type of language sample used by Carroll was obtained from the field of literature, and the individuals who produced the writing probably represented a highly selected group of verbally skilled individuals. In this study the verbal output of 'normal' Iowa school children, who are comparatively unskilled in language arts, was the object of investigation; (3) great differences in age, intelligence and environmental background undoubtedly existed between the two groups of subjects who produced the language studied in the two investigations. There is the possibility that, in general, the equation presented by Carroll will hold for some types of language samples but not for others.

On the other hand, there may be some question as to whether the notion involved in this equation is too narrow for psychological utility. Although an equation with only one parameter may be found to be adequate to describe the relationship between D and N, it would appear to be highly unlikely that the rate at which new words are added in connected discourse is so simply explained. The factor of number of topics, for example, has already been discussed as a possible factor influencing the rate at which new words are added, and probably other factors are also to be considered.

Attempts to fit a curve to represent the relationship between D and N reveals

that the function is of no simple nature. The study of the relationship between frequency of occurrence of a word and its rank demonstrates the important fact that the more frequently occurring words do not fit the same general function as do the rarer words. Considerations of the language structure suggests that perhaps it may be feasible to divide the words of any language sample into two categories. First, there are the interstitial or structural words which form the core of framework structure of language. Since these words carry little meaning beyond the verbal context in which they appear they may be termed intensional words in contrast to the second type of words, the content and action words that have, directly or indirectly, an extensional reference. Since the more frequently occurring words are of the intensional type, we have a basis for making an analysis of the language into two parts. In view of the fact that the extensional words serve to represent or are symbolic of the interests, attitudes, ambitions, etc. of the writer they would appear to be of greater psychological interest. There is some suggestion that if such a division of words could be made, a hyperbolic equation could be fitted to each division of words in the sample. One of the vitiating factors in attempts to find a lawful relationship between D and N may well lie in the difference between the ways in which these two types of words reach their maximum. The intensional words appear to reach a maximum in terms of D very rapidly, while the extensional words rise comparatively much more slowly and reach a maximum at a much later point on the curve.

VI. SUMMARY AND CONCLUSIONS

Three-thousand-word written language samples were obtained from 108 Iowa school children who had been selected to fill the cells of a factorial design which consisted of three levels each of I.Q., C.A., and locality (city, town, rural) as well as two equal groups of boys and girls. The subjects were asked to write about whatever they wanted to write about and in a free-writing situation for a short time each day until they had reached their quota of 3,000 words.

Each language sample was edited and the words tabulated according to a set of predetermined rules. From these tabulations, 20 language measures were obtained for each sample, and individual cumulative type-frequency curves were computed for a selected group of subjects. Language measures obtained were: (1) 100-word type-token ratio, (2) 500-word type-token ratio, (3) 1,000-word type-token ratio, (4) 3,000-word type-token ratio; number of tokens for (5) nounal, (6) verbal, (7) adjectival, (8) adverbal categories; number of types for (9) nounal, (10) verbal, (11) adjectival, (12) adverbal categories; type-token ratio for (13) nouns, (14) verbs, (15) adjectives, (16) adverbs; percentage of (17) nounal, (18) verbal, (19) adjectival, (20) adverbal types of the total types of these four parts of speech categories.

These data were analyzed in three ways in order to determine (a) the reliability of type-token ratios; (b) the ability of these measures to differentiate groups of individuals classified according to levels of I.Q., C.A., locality and sex; and (c) the mathematical relationship, if any, between the number of different words (D) and the size of the sample (N).

On the basis of these analyses the following conclusions can be drawn:

1. Segmental type-token ratios derived from samples of 3,000 words are highly reliable in (a) the agreement between

two independent sets of operations used to arrive at the numerical value of the type-token ratio and in (b) the relative constancy of the type-token ratio over a short span of time (about a week). A further implication, indicated by the fact that the reliability coefficient is a positive function of the size of sample, is that, in general, type-token ratios computed from a sample of 1,000 words in length are, for practical purposes, as reliable as are those computed from samples of 3,000 words in length and should prove satisfactory in all instances except when a high degree of precision is needed.

2. The results of the analyses of variance of the 20 language measures may be summarized briefly in the statement that the implication of these results is that the language measures employed can be used to characterize groups classified according to I.Q., C.A., locality and possibly sex, although the results for sex are practically negative. On the whole, the more highly developed the individual in terms of intelligence and age, the more highly differentiated his language structure appears to be. This is shown particularly by the type-token ratios. It may also be said that high I.Q. groups are characterized by the use of a proportionately greater number of nouns while the low I.Q. groups are characterized by the use of a greater percentage of verbs and adverbs.

3. An equation presented by Carroll was discussed and tests of its adequacy to describe these data were carried out. Its generality to these data was not substantiated, although a more general form of the same equation was found to give a fairly good fit. The relationship of the number of different words to the size of sample was found to be a complex one. Empirically fitted equations to represent these data are of such a character as to make prediction beyond the limits of the data hazardous.

BIBLIOGRAPHY

1. Boder, D. P. The adjective-verb-quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, *3*, 310-343.
2. Busemann, A. *Die Sprache der Jugend als Ausdruck der Eutwicklungsrhythmik*. Jena: Fischer, 1925, pp. x + 98.
3. Carroll, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.*, 1938, 2, 379-386.
4. Curme, G. O. *Parts of Speech and Accidence*. Boston: D. C. Heath and Co., 1935.
5. Fairbanks, H. Studies in language behavior: II. The quantitative differentiation of samples of spoken language. *Psychol. Monogr.*, 1944, *56*, 19-38.
6. Fossum, E. C. An analysis of certain characteristics of oral language of junior college students. Univ of Iowa Ph.D. Dissertation. 1941.
7. Johnson, W. Studies in Language Behavior: I. Objectives and methods. *Psychol. Monogr.*, 1944, *56*, 1-15.
8. Kent, G. H., & Rosanoff, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, *67*, 57-96.
9. Mann, M. B. Studies in Language Behavior: III. The quantitative differentiation of samples of written language. *Psychol. Monogr.*, 1944, *56*, 41-74.
10. Sanford, F. H. Speech and personality. *Psychol. Bull.*, 1942, *39*, 811-845.
11. Sanford, F. H. Speech and personality: a comparative case study. *Character & Pers.*, 1942, *19*, 169-198.
12. Skinner, B. F. The verbal summator and a method for the study of latent speech. *J. Psychol.*, 1936, 2, 71-107.
13. Skinner, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, *1*, 71-76.
14. Thorndike, R. L. On the number of words of any frequency of use. *Psychol. Rec.*, 1937, *1*, 379-406.
15. Zipf, G. K. *The Psychobiology of Language*. Boston: Houghton Mifflin Co., 1935, pp. ix + 336.

ovem-

unce-
f the

each

The
osely

lish
pear

The

ears
ame

to

P. 5, c
P. 7, c

P. 10,
P. 11,
P. 14,
thi

P. 15
tal
P. 18
lin
P. 2
lin
P. 22
P. 2
P. 2
P. 2

T
Dr. 7
is no

Psychological review.

VOLUME 56 WHOLE NO. 256
1944

ERRATA

P. 5, col. 1, line 15	*idfficulty* should be *difficulty*.
P. 7, col. 1, line 8	The following sentence should have appeared here as a footnote: A technique used by Dr. Hudson Jost, Ill. Inst. of Juvenile Research, Chicago, Ill.
P. 10, col. 1, line 20	*for* should be *on*.
P. 11, col. 2, line 20	*divising* should be *devising*.
P. 14, Insert at end of third paragraph	(Second Sample Made) Face is reddening. She chews her lips, then her fingers. She is breathing fast. Rubs her arms with palms of hands. Angry tone toward observer and flash-box is beginning to lessen. Voice is softer. Tears in eyes. (Third Sample Made) Face very red. Voice has become soft. Manner is apologetic. She is still trying. Lips are trembling. Voice a whisper, hesitating. Face has frightened expression. She mumbles, "I don't know, I don't know," and "I can't get it." Lips trembling greatly. She frowns, at point of crying aloud. "I can't get it. Oh, I never do any good!" Departs without a word. Silently shuts the door.
P. 15, Footnote to table	*socre* should be *score*.
P. 18, col. 2, fourth line from bottom	*groups* should be *group*.
P. 21, col. 2, seventh line from bottom	*his* should be *this*.
P. 22, col. 2, line 18	word *of* should be omitted.
P. 26, col. 1, line 21	*measured* should be *measure*.
P. 27, col. 1, line 16	*who* should be *which*.
P. 27, col. 2, line 8	Insert *test* after intelligence.

The Publications Office regrets that it is necessary to add this errata slip to Dr. Zander's Monograph. That the corrections to the text are made in this manner is not the responsibility of the author.

raphs

ration

ON, INC.

OIS